



The Geometry of Bayesian Programming

Ugo Dal Lago, Naohiko Hoshino

► To cite this version:

Ugo Dal Lago, Naohiko Hoshino. The Geometry of Bayesian Programming. LICS 2019 - 34th Annual ACM/IEEE Symposium on Logic in Computer Science, Jun 2019, Vancouver, Canada. pp.1-13, 10.1109/LICS.2019.8785663 . hal-02399343

HAL Id: hal-02399343

<https://hal.science/hal-02399343>

Submitted on 7 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Geometry of Bayesian Programming

Ugo Dal Lago*

Naohiko Hoshino†

January 7, 2020

Abstract

We give a geometry of interaction model for a typed λ -calculus endowed with operators for sampling from a continuous uniform distribution and soft conditioning, namely a paradigmatic calculus for higher-order Bayesian programming. The model is based on the category of measurable spaces and partial measurable functions, and is proved adequate with respect to both a distribution-based and a sampling-based operational semantics.

1 Introduction

Randomisation provides the most efficient algorithmic solutions, at least concretely, in many different contexts. A typical example is the one of primality testing, where the Miller-Rabin test [1, 2] remains the preferred choice despite polynomial time deterministic algorithms are available from many years now [3]. Probability theory can be exploited even more fundamentally in programming, by way of so-called probabilistic (or, more specifically, Bayesian) programming, as popularized by languages like, among others, **ANGLICAN** [4] or **CHURCH** [5]. This has stimulated research about probabilistic programming languages and their semantics [6, 7, 8], together with type systems [9, 10], equivalence methodologies [11, 12], and verification techniques [13].

Giving a satisfactory denotational semantics to higher-order functional languages is already problematic in presence of probabilistic choice [6, 14], and becomes even more challenging when continuous distributions and scoring are present. Recently, quasi-Borel spaces [15] have been proposed as a way to give semantics to calculi with all these features, and only very recently [16] this framework has been shown to be adaptable to a fully-fledged calculus for probabilistic programming, in which continuous distributions *and* soft-conditioning are present. Probabilistic coherent spaces [17] are fully abstract [8] for λ -calculi with discrete probabilistic choice, and can, with some effort, be adapted to calculi with sampling from continuous distributions [18], although without scoring.

A research path which has been studied only marginally, so far, consists in giving semantics to Bayesian higher-order programming languages through *interactive forms* of semantics, e.g. game semantics [19, 20] or the geometry of interaction [21]. One of the very first models for higher-order calculi with discrete probabilistic choice was in fact a game model, proved fully abstract for a probabilistic calculus with global ground references [7]. After more than ten years, a parallel form of Geometry of Interaction (GoI) and some game models have been introduced for λ -calculi with probabilistic choice [22, 23, 24], but in all these cases only discrete probabilistic choice can be handled, with the exception of a recent work on concurrent games and continuous distributions [25].

In this paper, we will report on some results about GoI models of higher-order Bayesian languages. The distinguishing features of the introduced GoI model can be summarised as follows:

- **Simplicity.** The category on which the model is defined is the one of measurable spaces and *partial* measurable functions, so it is completely standard from a measure-theoretic perspective.

*University of Bologna & INRIA

†Kyoto University

- **Expressivity.** As is well-known, the GoI construction [26, 27] allows to give semantics to calculi featuring higher-order functions and recursion. Indeed, our GoI model can be proved adequate for **PCFSS**, a fully-fledged calculus for probabilistic programming.
- **Flexibility.** The model we present is quite flexible, in the sense of being able to reflect the operational behaviour of programs as captured by *both* the distribution-based and the sampling-based semantics.
- **Intuitiveness.** GoI visualises the structure of programs in terms of graphs, from which dependencies between subprograms can be analyzed. Adequacy of our model provides diagrammatic reasoning principle about observational equivalence of **PCFSS**.

This paper’s contributions, beside the model’s definition, are two adequacy results which precisely relate our GoI model to the operational semantics, as expressed (following [28]), in both the distribution and sampling styles. As a corollary of our adequacy results, we show that integrating over the sampling-based operational semantics, one obtains precisely the distribution-based operational semantics.

1.1 Turning Measurable Spaces into a GoI Model

Before entering into the details of our model, it is worthwhile to give some hints about how the proposed model is obtained, and why it differs from similar GoI models from the literature.

The thread of work the proposed model stems from is the one of so-called memoryful geometry of interaction [29, 30]. The underlying idea of this paper is precisely the same: program execution is modelled as an interaction between the program and its environment, and memoisation takes place inside the program as a result of the interaction.

In the previous work on memoryful GoI by the second author with Hasuo and Muroya, the goal consisted in modelling a λ -calculus with *algebraic* effects. Starting from a monad together with some algebraic effects, they gave an adequate GoI model for such a calculus, which is applicable to a wide range of algebraic effects. In principle, then, their recipe could be applicable to **PCFSS**, since sampling-based operational semantics enables us to see scoring and sampling as algebraic effects acting on global states. However, that would not work for **PCFSS**, since the category **Meas** of measurable spaces¹ is not cartesian closed, and we thus *cannot* define a state monad by way of the exponential $S \Rightarrow S \times (-)$.

In this paper, we side step this issue by a series of translations, to be described in Section 4 below. Instead of looking for a state monad on **Meas**, we embed **Meas** into the category **Mealy** of **Int**-objects and Mealy machines (Section 5) and use a state monad *on this* category. This is doable because **Mealy** is a *compact closed category* given by the **Int**-construction [27]. The use of such compact closed categories (or, more generally, of traced monoidal categories) is the way GoI models higher-order functions.

1.2 Outline

The rest of the paper is organised as follows. After giving some necessary measure-theoretic preliminaries in Section 2 below, we introduce in Section 3 the language **PCFSS**, together with the two kinds of operational semantics we were referring to above. In Section 4, we introduce our GoI model informally, while in Section 5 a more rigorous treatment of the involved concepts is given, together with the adequacy results. We discuss in Section 10 an alternative way of giving a GoI semantics to **PCFSS** based on s-finite kernels, and we conclude in Section 12.

2 Measure-Theoretic Preliminaries

We recall some basic notions in measure theory that will be needed in the following. We also fix some useful notations. For more about measure theory, see standard textbooks such as [31].

¹We need to work on **Meas** because we want to give adequacy for distribution-based semantics.

A σ -algebra on a set X is a family Σ consisting of subsets of X such that $\emptyset \in \Sigma$; and if $A \in \Sigma$, then the complement $X \setminus A$ is in Σ ; and for any family $\{A_n \in \Sigma\}_{n \in \mathbb{N}}$, the intersection $\bigcap_{n \in \mathbb{N}} A_n$ is in Σ . A *measurable space* X is a set $|X|$ equipped with a σ -algebra Σ_X on $|X|$. We often confuse a measurable space X with its underlying set $|X|$. For example, we simply write $x \in X$ instead of $x \in |X|$. For measurable spaces X and Y , we say that a partial function $f: X \rightarrow Y$ (in this paper, we use \rightarrow for both partial functions and total functions) is *measurable* when for all $A \in \Sigma_Y$, the inverse image

$$\{x \in X : f(x) \text{ is defined and is equal to an element of } A\}$$

is in Σ_X . A *measurable function* from X to Y is a totally defined partial measurable function. A (partial) measurable function $f: X \rightarrow Y$ is *invertible* when there is a measurable function $g: Y \rightarrow X$ such that $g \circ f$ and $f \circ g$ are identities. In this case, we say that f is an isomorphism from X to Y and say that X is isomorphic to Y .

We denote a singleton set $\{*\}$ by 1, and we regard the latter as a measurable space by endowing it with the trivial σ -algebra. We also regard the empty set \emptyset as a measurable space in the obvious way. In this paper, \mathbb{N} denotes the measurable set of all non-negative integers equipped with the σ -algebra consisting of all subsets of \mathbb{N} , and \mathbb{R} denotes the measurable set of all real numbers equipped with the σ -algebra consisting of *Borel sets*, that is, the least σ -algebra that contains all open subsets of \mathbb{R} . By the definition of $\Sigma_{\mathbb{R}}$, a function $f: \mathbb{R} \rightarrow \mathbb{R}$ is measurable whenever $f^{-1}(U) \in \Sigma_{\mathbb{R}}$ for all open subsets $U \subseteq \mathbb{R}$. Therefore, all continuous functions on \mathbb{R} are measurable.

When Y is a subset of the underlying set of a measurable space X , we can equip Y with a σ -algebra $\Sigma_Y = \{A \cap Y : A \in \Sigma_X\}$. This way, we regard the unit interval and the set of all non-negative real numbers as measurable spaces, and indicate them as follows:

$$\mathbb{R}_{[0,1]} = \{a \in \mathbb{R} : 0 \leq a \leq 1\}, \quad \mathbb{R}_{\geq 0} = \{a \in \mathbb{R} : a \geq 0\}$$

For measurable spaces X and Y , we define the product measurable space $X \times Y$ and the coproduct measurable space $X + Y$ by

$$\begin{aligned} |X \times Y| &= |X| \times |Y|, \\ |X + Y| &= \{(\bullet, x) : x \in X\} \cup \{(\circ, y) : y \in Y\} \end{aligned}$$

where the underlying σ -algebras are:

$$\begin{aligned} \Sigma_{X \times Y} &= \text{the least } \sigma\text{-algebra such that } A \times B \in \Sigma_{X \times Y} \\ &\quad \text{for all } A \in \Sigma_X \text{ and } B \in \Sigma_Y, \\ \Sigma_{X+Y} &= \{\{\bullet\} \times A \cup \{\circ\} \times B : A \in \Sigma_X \text{ and } B \in \Sigma_Y\}. \end{aligned}$$

We assume that \times has higher precedence than $+$, i.e., we write $X + Y \times Z$ for $X + (Y \times Z)$. In this paper, we always regard finite products \mathbb{R}^n as the product measurable space on \mathbb{R} . It is well-known that the σ -algebra $\Sigma_{\mathbb{R}^n}$ is the set of all Borel sets, i.e., $\Sigma_{\mathbb{R}^n}$ is the least one that contains all open subsets of \mathbb{R}^n . Partial measurable functions are closed under compositions, products and coproducts.

Let X be a measurable space. A *measure* μ on X is a function from Σ_X to $[0, \infty]$ that is the set of all non-negative real numbers extended with ∞ , such that

- $\mu(\emptyset) = 0$; and
- for any mutually disjoint family $\{A_n \in \Sigma_X\}_{n \in \mathbb{N}}$, we have $\sum_{n \in \mathbb{N}} \mu(A_n) = \mu(\bigcup_{n \in \mathbb{N}} A_n)$.

We say that a measure μ on X is *finite* when $\mu(X) < \infty$ and that it is *σ -finite* if $X = \bigcup_{n \in \mathbb{N}} X_n$ for some family $\{X_n \in \Sigma_X\}_{n \in \mathbb{N}}$ satisfying $\mu(X_n) < \infty$.

For a measurable space X , we write \emptyset_X for a measure on X given by $\emptyset_X(A) = 0$ for all $A \in \Sigma_X$. If μ is a measure on a measurable space X , then for any non-negative real number a , the function $(a\mu)(A) = a(\mu(A))$ is also a measure on X . The *Borel measure* μ_{Borel} on \mathbb{R}^n is the unique measure that satisfies

$$\mu_{\text{Borel}}([a_1, b_1] \times \cdots \times [a_n, b_n]) = \prod_{1 \leq i \leq n} |a_i - b_i|.$$

We define the Borel measure μ_{Borel} on 1 by $\mu_{\text{Borel}}(1) = 1$. For a measurable function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and a measurable subset $X \subseteq \mathbb{R}^n$, we denote the integral of f with respect to the Borel measure restricted to X by

$$\int_X f(u) \, du.$$

For a measurable space X and for an element $x \in X$, a *Dirac measure* δ_x on X is given by

$$\delta_x(A) = [x \in A] = \begin{cases} 1, & \text{if } x \in A; \\ 0, & \text{if } x \notin A. \end{cases}$$

The square bracket notation in the right hand side is called Iverson's bracket. In general, for a proposition P , we have $[P] = 1$ when P is true and $[P] = 0$ when P is false.

Proposition 2.1. *For every σ -finite measures μ on a measurable space X and ν on a measurable space Y , there is a unique measure $\mu \times \nu$ on $X \times Y$ such that $(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$ for all $A \in \Sigma_X$ and $B \in \Sigma_Y$.*

The measure $\mu \times \nu$ is called the *product measure* of μ and ν . For example, the Borel measure on \mathbb{R}^2 is the product measure of the Borel measure on \mathbb{R} .

Finally, let us recall the notion of a kernel, which is a well-known concept in the theory of stochastic processes. For measurable spaces X and Y , a *kernel* from X to Y is a function $k: X \times \Sigma_Y \rightarrow [0, \infty]$ such that for any $x \in X$, the function $k(x, -)$ is a measure on Y , and for any $A \in \Sigma_Y$, the function $k(-, A)$ is measurable. Notions of *finite* and *σ -finite* kernels can be naturally given, following the eponymous constraints on measures. Those kernels which can be expressed as the sum of countably many finite kernels are said to be *s-finite* [32]. We use kernels to give semantics for our probabilistic programming language, to be defined in the next section.

3 Syntax and Operational Semantics

3.1 Syntax and Type System

Our language **PCFSS** for higher order Bayesian programming can be seen as Plotkin's **PCF** endowed with real numbers, measurable functions, sampling from the uniform distribution on $\mathbb{R}_{[0,1]}$ and soft-conditioning. We first define types A, B, \dots , values V, W, \dots and terms M, N, \dots as follows:

$$\begin{aligned} A, B ::= & \text{Unit} \mid \text{Real} \mid A \rightarrow B, \\ V, W ::= & \text{skip} \mid x \mid \lambda x^A. M \mid r_a \mid \text{fix}_{A,B}(f, x, M), \\ M, N ::= & V \mid VW \mid \text{let } x \text{ be } M \text{ in } N \mid \text{ifz}(V, M, N) \\ & \mid F(V_1, \dots, V_{|F|}) \mid \text{sample} \mid \text{score}(V). \end{aligned}$$

Here, x varies over a countably infinite set of variable symbols, and a varies over the set \mathbb{R} of all real numbers. Each *function identifier* F is associated with a measurable function fun_F from $\mathbb{R}^{|F|}$ to \mathbb{R} . For terms M and N , we write $M\{N/x\}$ for the capture-avoiding substitution of x in M by N .

Terms in **PCFSS** are restricted to be *A-normal forms*, in order to make some of the arguments about our semantics simpler. This restriction is harmless to the language expressive power, thanks to the presence of **let**-bindings. For example, term application MN can be defined to be **let** x **be** M **in** **let** y **be** N **in** xy .

The term constructor **score** and the constant **sample** enable probabilistic programming in **PCFSS**. Evaluation of **score**(r_a) has the effect of multiplying the weight of the current probabilistic branch by $|a|$, this way enabling a form of *soft-conditioning*. The constant **sample** generates

| | | |
|--|---|--|
| $\frac{x : A \in \Delta}{\Delta \vdash x : A}$ | $\frac{a \in \mathbb{R}}{\Delta \vdash r_a : \mathbf{Real}}$ | $\frac{\Delta \vdash V_i : \mathbf{Real} \text{ for all } i \leq F }{\Delta \vdash F(V_1, \dots, V_{ F }) : \mathbf{Real}}$ |
| $\frac{\Delta \vdash V : A \rightarrow B \quad \Delta \vdash W : A}{\Delta \vdash VW : B}$ | $\frac{\Delta \vdash M : B \quad \Delta, x : B \vdash N : A}{\Delta \vdash \text{let } x \text{ be } M \text{ in } N : A}$ | |
| $\frac{\Delta, x : A \vdash M : B}{\Delta \vdash \lambda x^A. M : A \rightarrow B}$ | $\frac{\Delta, f : A \rightarrow B, x : A \vdash M : B}{\Delta \vdash \text{fix}_{A,B}(f, x, M) : A \rightarrow B}$ | |
| $\frac{}{\Delta \vdash \text{skip} : \mathbf{Unit}}$ | $\frac{\Delta \vdash V : \mathbf{Real} \quad \Delta \vdash M : A \quad \Delta \vdash N : A}{\Delta \vdash \text{ifz}(V, M, N) : A}$ | |
| $\frac{}{\Delta \vdash \text{sample} : \mathbf{Real}}$ | $\frac{\Delta \vdash V : \mathbf{Real}}{\Delta \vdash \text{score}(V) : \mathbf{Unit}}$ | |

Figure 1: Typing Rules

a real number randomly drawn from the uniform distribution on $\mathbb{R}_{[0,1]}$. Only one sampling mechanism is sufficient because we can model sampling from other standard distributions by composing **sample** with measurable functions [33].

Terms can be typed in a natural way. A *context* Δ is a finite sequence consisting of pairs of a variable and a type such that every variable appears in Δ at most once. A *type judgement* is a triple $\Delta \vdash M : A$ consisting of a context Δ , a term M and a type A . We say that a type judgement $\Delta \vdash M : A$ is *derivable* when we can derive $\Delta \vdash M : A$ from the typing rules in Figure 1. Here, the type of **sample** is **Real**, and the type of **score**(V) is **Unit** because **sample** returns a real number, and the purpose of scoring is its side effect.

In the sequel, we only consider derivable type judgements and *typable* closed terms, that is, closed terms M such that $\vdash M : A$ is derivable for some type A .

3.2 Distribution-Based Operational Semantics

We define distribution-based operational semantics following [28] where, however, a σ -algebra on the set of terms is necessary so as to define evaluation results of terms to be distributions (i.e. measures) over values. In this paper, we only consider evaluation of terms of type **Real** and avoid introducing σ -algebras on sets of closed terms, thus greatly simplifying the overall development.

Distribution-based operational semantics is a function that sends a closed term $M : \mathbf{Real}$ to a measure μ on \mathbb{R} . Because of the presence of **score**, the measure may not be a probabilistic measure, i.e., $\mu(\mathbb{R})$ may be larger than 1, but the idea of distribution-based operational semantics is precisely that of associating each closed term of type **Real** with a measure over \mathbb{R} .

As common in call-by-value programming languages, evaluation is defined by way of *evaluation contexts*:

$$E[-] ::= [-] \mid \text{let } x \text{ be } E[-] \text{ in } M.$$

The distribution-based operational semantics of **PCFSS** is a family of binary relations $\{\Rightarrow_n\}_{n \in \mathbb{N}}$ between closed terms of type **Real** and measures on \mathbb{R} inductively defined by the evaluation rules in Figure 2 where the evaluation rule for **score** is inspired from the one in [32]. The binary relation $\xrightarrow{\text{red}}$ in the precondition of the third rule in Figure 2 is called *deterministic reduction* and is defined

as follows as a relation on closed terms:

$$\begin{aligned}
& (\lambda \mathbf{x}^A. \mathbf{M}) \mathbf{V} \xrightarrow{\text{red}} \mathbf{M}\{\mathbf{V}/\mathbf{x}\}, \\
& \text{let } \mathbf{x} \text{ be } \mathbf{V} \text{ in } \mathbf{M} \xrightarrow{\text{red}} \mathbf{M}\{\mathbf{V}/\mathbf{x}\}, \\
& \text{fix}_{A,B}(\mathbf{f}, \mathbf{x}, \mathbf{M}) \mathbf{V} \xrightarrow{\text{red}} \mathbf{M}\{\text{fix}_{A,B}(\mathbf{f}, \mathbf{x}, \mathbf{M})/\mathbf{f}, \mathbf{V}/\mathbf{x}\}, \\
& \text{ifz}(\mathbf{r}_a, \mathbf{M}, \mathbf{N}) \xrightarrow{\text{red}} \begin{cases} \mathbf{M}, & \text{if } a = 0, \\ \mathbf{N}, & \text{if } a \neq 0, \end{cases} \\
& \mathbf{F}(\mathbf{r}_a, \dots, \mathbf{r}_b) \xrightarrow{\text{red}} \mathbf{r}_{\text{funF}(a, \dots, b)}.
\end{aligned}$$

The last evaluation rule in Figure 2 makes sense because k in the precondition is a kernel from $\mathbb{R}_{[0,1]}$ to \mathbb{R} :

Lemma 3.1. *For any $n \in \mathbb{N}$ and for any term*

$$\mathbf{x}_1 : \mathbf{Real}, \dots, \mathbf{x}_m : \mathbf{Real} \vdash \mathbf{M} : \mathbf{Real},$$

there is a finite kernel k from \mathbb{R}^m to \mathbb{R} such that for any $u \in \mathbb{R}^m$ and for any measure μ on \mathbb{R} ,

$$\mathbf{M}\{\mathbf{r}_{a_1}/\mathbf{x}_1, \dots, \mathbf{r}_{a_m}/\mathbf{x}_m\} \Rightarrow_n \mu \iff \mu = k(u, -)$$

where $u = (a_1, \dots, a_m)$.

Proof. Let Δ be a context of the form $\mathbf{x}_1 : \mathbf{Real}, \dots, \mathbf{x}_m : \mathbf{Real}$. In this proof, for a finite sequence $u = (a_1, \dots, a_m) \in \mathbb{R}^m$, and for a term $\Delta \vdash \mathbf{M} : \mathbf{A}$, we denote

$$\mathbf{M}\{\mathbf{r}_{a_1}/\mathbf{x}_1, \dots, \mathbf{r}_{a_m}/\mathbf{x}_m\}$$

by $\mathbf{M}\{\mathbf{r}_u/\Delta\}$. We prove the statement by induction on $n \in \mathbb{N}$. (Base case) Let k be the zero kernel from \mathbb{R}^m to \mathbb{R} :

$$k(u, A) = 0.$$

Then for any $u = (a_1, \dots, a_m) \in \mathbb{R}^m$,

$$\mathbf{M}\{\mathbf{r}_{a_1}/\mathbf{x}_1, \dots, \mathbf{r}_{a_m}/\mathbf{x}_m\} \Rightarrow_0 \mu \iff \mu = \emptyset_{\mathbb{R}} \iff \mu = k(u, -).$$

(Induction step) We define a *redex* \mathbf{R} by

$$\begin{aligned}
\mathbf{R} ::= & \text{score}(\mathbf{V}) \mid \text{sample} \mid (\lambda \mathbf{x}^A. \mathbf{M}) \mathbf{V} \mid \text{fix}_{A,B}(\mathbf{f}, \mathbf{x}, \mathbf{M}) \mathbf{V} \\
& \mid \mathbf{F}(\mathbf{V}, \dots, \mathbf{W}) \mid \text{let } \mathbf{x} \text{ be } \mathbf{V} \text{ in } \mathbf{M} \mid \text{ifz}(\mathbf{r}_a, \mathbf{M}, \mathbf{N}).
\end{aligned}$$

We note that \mathbf{V}, \mathbf{W} in the above BNF can be variables. By induction on the size of type derivation, we can show that every term $\Delta \vdash \mathbf{M} : \mathbf{A}$ is either a value or of the form $\mathbf{E}[\mathbf{R}]$ for some evaluation context $\mathbf{E}[-]$ and some redex \mathbf{R} . Given a term $\Delta \vdash \mathbf{M} : \mathbf{A}$ where $\Delta = \mathbf{x}_1 : \mathbf{Real}, \dots, \mathbf{x}_m : \mathbf{Real}$, we prove the induction step by case analysis.

- If $\Delta \vdash \mathbf{M} : \mathbf{Real}$ is a value, then \mathbf{M} is either a variable \mathbf{x}_i or a constant \mathbf{r}_a . When \mathbf{M} is a variable \mathbf{x}_i , we have

$$\mathbf{x}_i\{\mathbf{r}_{a_1}/\mathbf{x}_1, \dots, \mathbf{r}_{a_m}/\mathbf{x}_m\} \equiv \mathbf{r}_{a_i} \Rightarrow_{n+1} \mu \iff \mu = \delta_{a_i}.$$

When \mathbf{M} is a constant \mathbf{r}_a , we have

$$\mathbf{r}_a\{\mathbf{r}_{a_1}/\mathbf{x}_1, \dots, \mathbf{r}_{a_m}/\mathbf{x}_m\} \equiv \mathbf{r}_a \Rightarrow_{n+1} \mu \iff \mu = \delta_a.$$

Both $k, h : \mathbb{R}^m \times \Sigma_{\mathbb{R}} \rightarrow [0, \infty]$ given by

$$k((a_1, \dots, a_m), A) = \delta_{a_i}(A), \quad h((a_1, \dots, a_m), A) = \delta_a(A)$$

are finite kernels from \mathbb{R}^m to \mathbb{R} .

- If $\Delta \vdash M : \mathbf{Real}$ is of the form $E[\mathbf{sample}]$, then by induction hypothesis, there is a finite kernel from \mathbb{R}^{m+1} to \mathbb{R} such that for any $u \in \mathbb{R}^{m+1}$,

$$E[y]\{\mathbf{r}_u / (\Delta, y : \mathbf{Real})\} \Rightarrow_n \mu \iff \mu = k(u, -).$$

We define a kernel h from \mathbb{R}^m to \mathbb{R} by

$$h((a_1, \dots, a_m), A) = \int_{\mathbb{R}_{[0,1]}} k((a_1, \dots, a_m, a), A) da.$$

This is a finite kernel because for any non-negative measurable function $f: \mathbb{R} \times \dots \times \mathbb{R} \rightarrow \mathbb{R}$ bounded by $d > 0$,

$$(b, \dots, c) \mapsto \int_{\mathbb{R}_{[0,1]}} f(a, b, \dots, c) da$$

is measurable and bounded by d . See [31, Theorem 18.3]. Then, for any $u = (a_1, \dots, a_m) \in \mathbb{R}^m$,

$$\begin{aligned} E[\mathbf{sample}]\{\mathbf{r}_u / \Delta\} \Rightarrow_{n+1} \mu &\iff \mu = \int_{\mathbb{R}_{[0,1]}} k((a_1, \dots, a_m, a), -) da \\ &\iff \mu = h(u, -). \end{aligned}$$

- If $\Delta \vdash M : B$ is of the form $E[\mathbf{score}(\mathbf{x}_i)]$ for some $i \in \{1, 2, \dots, m\}$, then by induction hypothesis, there is a finite kernel k from \mathbb{R}^m to \mathbb{R} such that for any $u \in \mathbb{R}^m$,

$$E[\mathbf{skip}]\{\mathbf{r}_u / \Delta\} \Rightarrow_n \mu \iff \mu = k(u, -).$$

We define a finite kernel $h: \mathbb{R}^m$ to \mathbb{R} by

$$h((a_1, \dots, a_m), A) = |a_i| k((a_1, \dots, a_m), A).$$

Then, for any $u = (a_1, \dots, a_m) \in \mathbb{R}^m$,

$$\begin{aligned} E[\mathbf{score}(\mathbf{x}_i)]\{\mathbf{r}_u / \Delta\} \Rightarrow_{n+1} \mu &\iff E[\mathbf{skip}]\{\mathbf{r}_u / \Delta\} \Rightarrow_n \nu \text{ and } \mu = |a_i| \nu \\ &\iff \mu = h(u, -). \end{aligned}$$

- If $\Delta \vdash M : B$ is of the form $E[\mathbf{score}(\mathbf{r}_a)]$ for some $a \in \mathbb{R}$, then by induction hypothesis, there is a finite kernel k from \mathbb{R}^m to \mathbb{R} such that for any $u \in \mathbb{R}^m$,

$$E[\mathbf{skip}]\{\mathbf{r}_u / \Delta\} \Rightarrow_n \mu \iff \mu = k(u, -).$$

We define a finite kernel $h: \mathbb{R}^m$ to \mathbb{R} by

$$h((a_1, \dots, a_m), A) = |a| k((a_1, \dots, a_m), A).$$

Then, for any $u = (a_1, \dots, a_m) \in \mathbb{R}^m$,

$$\begin{aligned} E[\mathbf{score}(\mathbf{x}_i)]\{\mathbf{r}_u / \Delta\} \Rightarrow_{n+1} \mu &\iff E[\mathbf{skip}]\{\mathbf{r}_u / \Delta\} \Rightarrow_n \nu \text{ and } \mu = |a| \nu \\ &\iff \mu = h(u, -). \end{aligned}$$

- If $\Delta \vdash M : B$ is of the form $E[(\lambda \mathbf{x}^A. N) V]$, then by induction hypothesis, there is a finite kernel k from \mathbb{R}^m to \mathbb{R} such that for all $u \in \mathbb{R}^m$,

$$E[N \{V/\mathbf{x}\}]\{\mathbf{r}_u / \Delta\} \Rightarrow_n \mu \iff \mu = k(u, -).$$

Hence,

$$\begin{aligned} E[(\lambda \mathbf{x}^A. N) V]\{\mathbf{r}_u / \Delta\} \Rightarrow_{n+1} \mu &\iff E[N \{V/\mathbf{x}\}]\{\mathbf{r}_u / \Delta\} \Rightarrow_n \mu \\ &\iff \mu = k(u, -). \end{aligned}$$

- If $\Delta \vdash M : B$ is of the form $E[\text{fix}_{A,B}(f, x, N) V]$, then by induction hypothesis, there is a finite kernel k from \mathbb{R}^m to \mathbb{R} such that for all $u \in \mathbb{R}^m$,

$$E[N\{\text{fix}_{A,B}(f, x, N)/f, V/x\}]\{r_u/\Delta\} \Rightarrow_n \mu \iff \mu = k(u, -).$$

Hence,

$$\begin{aligned} E[\text{fix}_{A,B}(f, x, N) V] \Rightarrow_{n+1} \mu &\iff E[N\{\text{fix}_{A,B}(f, x, N)/f, V/x\}]\{r_u/\Delta\} \Rightarrow_n \mu \\ &\iff \mu = k(u, -). \end{aligned}$$

- If $\Delta \vdash M : \text{Real}$ is of the form $E[F(V_1, \dots, V_{|F|})]$, then V_i is equal to either a variable or a constant r_a . For simplicity, we suppose that $|F| = 2$ and $V_1 = x_i$ and $V_2 = r_a$. By induction hypothesis, there is a finite kernel from \mathbb{R}^{m+1} to \mathbb{R} such that for all $u \in \mathbb{R}^{m+1}$,

$$E[y]\{r_u/(\Delta, y : \text{Real})\} \Rightarrow_n \mu \iff \mu = k(u, -).$$

We define a finite kernel h from \mathbb{R}^m to \mathbb{R} by

$$h((a_1, \dots, a_m), A) = k((a_1, \dots, a_m, \text{fun}_F(a_i, a)), A).$$

Then, for any $u = (a_1, \dots, a_m) \in \mathbb{R}^m$,

$$\begin{aligned} E[F[x_i, r_a]]\{r_u/\Delta\} \Rightarrow_{n+1} \mu &\iff E[y]\{r_u/\Delta, r_{\text{fun}_F(a_i, a)}/y\} \Rightarrow_n \mu \\ &\iff \mu = k((u, \text{fun}_F(a_i, a)), -) = h(u, -). \end{aligned}$$

- If $\Delta \vdash M : \text{Real}$ is of the form $\text{let } x \text{ be } V \text{ in } N$, then by induction hypothesis, there is a finite kernel k from \mathbb{R}^m to \mathbb{R} such that for all $u \in \mathbb{R}^m$,

$$E[N\{V/x\}]\{r_u/\Delta\} \Rightarrow_n \mu \iff \mu = k(u, -).$$

Hence,

$$\begin{aligned} E[\text{let } x \text{ be } V \text{ in } N]\{r_u/\Delta\} \Rightarrow_{n+1} \mu &\iff E[N\{V/x\}]\{r_u/\Delta\} \Rightarrow_n \mu \\ &\iff \mu = k(u, -). \end{aligned}$$

- If $\Delta \vdash M : \text{Real}$ is of the form $E[\text{ifz}(x_i, N, L)]$ for some $i \in \{1, 2, \dots, m\}$, then by induction hypothesis, there are finite kernels k and k' from \mathbb{R}^m to \mathbb{R} such that for any $u \in \mathbb{R}^m$,

$$\begin{aligned} E[N]\{r_u/\Delta\} \Rightarrow_n \mu &\iff \mu = k(u, -), \\ E[L]\{r_u/\Delta\} \Rightarrow_n \mu &\iff \mu = k'(u, -). \end{aligned}$$

We define a finite kernel h from \mathbb{R}^m to \mathbb{R} by

$$h(u, A) = \begin{cases} k(u, A), & \text{if } a_i = 0, \\ k'(u, A), & \text{if } a_i \neq 0 \end{cases} \quad \text{where } u = (a_1, \dots, a_m).$$

Then, for any $u \in \mathbb{R}^m$,

$$\begin{aligned} E[\text{ifz}(x_i, N, L)]\{r_u/\Delta\} \Rightarrow_{n+1} \mu &\iff (E[N]\{r_u/\Delta\} \Rightarrow_n \mu \text{ and } a_i = 0) \\ &\quad \text{or } (E[L]\{r_u/\Delta\} \Rightarrow_n \mu \text{ and } a_i \neq 0) \\ &\iff \mu = h(u, -). \end{aligned}$$

- If $\Delta \vdash M : \text{Real}$ is of the form $E[\text{ifz}(r_0, N, L)]$, then by induction hypothesis, there is a finite kernel k from \mathbb{R}^m to \mathbb{R} such that for any $u \in \mathbb{R}^m$,

$$E[N]\{r_u/\Delta\} \Rightarrow_n \mu \iff \mu = k(u, -).$$

Hence,

$$\begin{aligned} E[\text{ifz}(r_0, N, L)]\{r_u/\Delta\} \Rightarrow_{n+1} \mu &\iff E[N]\{r_u/\Delta\} \Rightarrow_n \mu \\ &\iff \mu = k(u, -). \end{aligned}$$

$$\boxed{
\begin{array}{c}
\frac{}{M \Rightarrow_0 \emptyset_{\mathbb{R}}} \quad \frac{n > 0}{r_a \Rightarrow_n \delta_a} \quad \frac{M \xrightarrow{\text{red}} N \quad E[N] \Rightarrow_n \mu}{E[M] \Rightarrow_{n+1} \mu} \\
\frac{E[\text{skip}] \Rightarrow_n \mu}{E[\text{score}(r_a)] \Rightarrow_{n+1} |a| \mu} \\
\frac{E[r_a] \Rightarrow_n k(a, -) \quad \text{for all } a \in \mathbb{R}_{[0,1]}}{E[\text{sample}] \Rightarrow_{n+1} \int_{\mathbb{R}_{[0,1]}} k(a, -) da}
\end{array}
}$$

Figure 2: Evaluation Rules of Distribution-Based Operational Semantics

- If $\Delta \vdash M : \mathbf{Real}$ is of the form $E[\text{ifz}(r_a, N, L)]$ for some real number $a \neq 0$, then by induction hypothesis, there is a finite kernel k from \mathbb{R}^m to \mathbb{R} such that

$$E[L]\{r_u/\Delta\} \Rightarrow_n \mu \iff \mu = k(u, -).$$

Hence,

$$\begin{aligned}
E[\text{ifz}(r_a, N, L)]\{r_u/\Delta\} \Rightarrow_{n+1} \mu &\iff E[L]\{r_u/\Delta\} \Rightarrow_n \mu \\
&\iff \mu = k(u, -).
\end{aligned}$$

□

Lemma 3.1 implies that the relations \Rightarrow_n can be seen as functions from the set of closed terms of type **Real** to the set of measures on \mathbb{R} .

The step-indexed distribution-based operational semantics approximates the evaluation of closed terms by restricting the number of reduction steps. Thus, the limit of the step-indexed distribution-based operational semantics represents the “true” result of evaluating the underlying term.

Definition 3.1. For a closed term $M : \mathbf{Real}$ and a measure μ on \mathbb{R} , we write $M \Rightarrow_{\infty} \mu$ when there is a family of measures $\{\mu_n\}_{n \in \mathbb{N}}$ on \mathbb{R} such that $M \Rightarrow_n \mu_n$ and for all $A \in \Sigma_{\mathbb{R}}$,

$$\mu(A) = \sup_{n \in \mathbb{N}} \mu_n(A).$$

The binary relation \Rightarrow_{∞} is a function from the set of closed terms of type **Real** to the set of measures on \mathbb{R} . This follows from Lemma 3.1 and that the family of measures $\{\mu_n\}_{n \in \mathbb{N}}$ on \mathbb{R} such that $M \Rightarrow_n \mu_n$ forms an ascending chain $\mu_0 \leq \mu_1 \leq \dots$ with respect to the pointwise order. Moreover, it can be proved that for any $\mathbf{x}_1 : \mathbf{Real}, \dots, \mathbf{x}_m : \mathbf{Real} \vdash M : \mathbf{Real}$, there is an s-finite kernel k given by $M\{r_{a_1}/\mathbf{x}_1, \dots, r_{a_m}/\mathbf{x}_m\} \Rightarrow_{\infty} k((a_1, \dots, a_m), -)$.

3.3 Sampling-Based Operational Semantics

PCFSS can be endowed with another form of operational semantics, closer in spirit to inference algorithms, called the *sampling-based* operational semantics. The way we formulate it is deeply inspired from the one in [28].

The idea behind sampling-based operational semantics is to give the evaluation result of each probabilistic branch somehow independently. We specify each probabilistic branch by two parameters: one is a sequence of random draws, which will be consumed by **sample**; the other is a likelihood measure called weight, which will be modified by **score**.

Definition 3.2. A configuration is a triple (M, a, u) consisting of a closed term $M : \mathbf{Real}$, a real number $a \geq 0$ called the configuration’s weight, and a finite sequence u of real numbers in $\mathbb{R}_{[0,1]}$, called its trace.

$$\boxed{
\begin{array}{c}
\frac{M \xrightarrow{\text{red}} N}{(M, b, u) \rightarrow (N, b, u)} \\
(E[\text{score}(\mathbf{r}_a)], b, u) \rightarrow (E[\text{skip}], |a| b, u) \\
(E[\text{sample}], a, b :: u) \rightarrow (E[\mathbf{r}_b], a, u)
\end{array}
}$$

Figure 3: Evaluation Rules of Sampling-Based Operational Semantics

Below, we write ε for the empty sequence. For a real number a and a finite sequence u consisting of real numbers, we write $a :: u$ for the finite sequence obtained by putting a on the head of u . In Figure 3, we give the evaluation rules of sampling-based operational semantics where $\xrightarrow{\text{red}}$ is the deterministic reduction relation introduced in the previous section. We denote the reflective transitive closure of \rightarrow by \rightarrow^* . Intuitively, $(M, 1, u) \rightarrow^* (\mathbf{r}_a, b, \varepsilon)$ means that by evaluating M , we get the real number a with weight b consuming *all* the random draws in u .

4 Towards Mealy Machine Semantics

In this section, we give some intuitions about our GoI model, which we also call *Mealy machine semantics*. Giving Mealy machine semantics for **PCFSS** requires translating **PCFSS** into the linear λ -calculus. This is because GoI is a semantics for linear logic, and is thus tailored for calculi in which terms are treated as resources. Schematically, Mealy machine semantics for **PCFSS** translates terms in **PCFSS** into Mealy machines in the following way.

$$\begin{array}{c}
\mathbf{PCFSS} \\
\downarrow (1) \text{ Moggi's translation} \\
\text{Moggi's meta-language} + \mathbf{sample} + \mathbf{score} \\
\downarrow (2) \text{ Girard translation} \\
\text{the linear } \lambda\text{-calculus} + \mathbf{sample} + \mathbf{score} \\
\downarrow (3) \\
\text{proof structures} + \mathbf{sample} + \mathbf{score} \\
\downarrow (4) \\
\text{Mealy machines .}
\end{array}$$

In Section 4.1, we explain the first three steps. The last step deserves to be explained in more detail, which we do in Section 4.2. For the sake of simplicity, we ignore the translation of conditional branching and the fixed point operator.

4.1 From PCFSS to Proof Structures

4.1.1 Moggi's Translation

In the first step, we translate **PCFSS** into an extension of the Moggi's meta-language by Moggi's translation [34]. Here, in order to translate scoring and sampling in **PCFSS**, we equip Moggi's meta-language with base types **Unit** and **Real** and the following terms:

$$\frac{a \in \mathbb{R}}{\Delta \vdash \mathbf{r}_a : \mathbf{Real}}, \quad \frac{\Delta \vdash M : \mathbf{Real}}{\text{score}(M) : \mathbf{TUnit}}, \quad \frac{}{\Delta \vdash \mathbf{sample} : \mathbf{TReal}}$$

where \mathbf{T} is the monad of Moggi's meta-language. Any type A of **PCFSS** is translated into the type A^\sharp defined as follows:

$$\mathbf{Unit}^\sharp = \mathbf{Unit}, \quad \mathbf{Real}^\sharp = \mathbf{Real}, \quad (A \rightarrow B)^\sharp = A^\sharp \rightarrow \mathbf{T}B^\sharp.$$

Terms **sample** and **score**($-$) in **PCFSS** are translated into **sample** and **score**($-$) in Moggi's meta-language respectively. See [34] for more detail about Moggi's translation.

4.1.2 Girard Translation

We next translate the extended Moggi's meta-language into an extension of the linear λ -calculus, by way of the so-called Girard translation [35]. Types are given by

$$A, B ::= \mathbf{Unit} \mid \mathbf{Real} \mid \mathbf{State} \mid A^\perp \mid A \otimes B \mid A \wp B \mid !A$$

where \mathbf{Unit} , \mathbf{Real} and \mathbf{State} are base types, and terms are generated by the standard term constructors of the linear λ -calculus, plus the following rules:

$$\frac{a \in \mathbb{R}}{\Delta \vdash r_a : \mathbf{Real}} \quad \frac{\Delta \vdash M : !\mathbf{Real}}{\Delta \vdash \mathbf{score}(M) : \mathbf{State} \multimap \mathbf{State} \otimes !\mathbf{Unit}} \quad \frac{}{\Delta \vdash \mathbf{sample} : \mathbf{State} \multimap \mathbf{State} \otimes !\mathbf{Real}}$$

(as customary in linear logic, $A \multimap B$ is an abbreviation of $A^\perp \wp B$). These typing rules are derived from the following translation $(-)^b$ of types of the extended Moggi's meta-language into types of the extended linear λ -calculus:

$$\mathbf{Unit}^b = \mathbf{Unit}, \quad \mathbf{Real}^b = \mathbf{Real}, \quad (A \rightarrow B)^b = !A^b \multimap B^b, \\ (\mathbf{TA})^b = \mathbf{State} \multimap \mathbf{State} \otimes !A^b$$

The definition of $(\mathbf{TA})^b$ is motivated by the following categorical observation: let \mathcal{L} be the syntactic category of the extended linear λ -calculus, which is a symmetric monoidal closed category endowed with a comonad $! : \mathcal{L} \rightarrow \mathcal{L}$ with certain coherence conditions (see e.g. [36]), and let $\mathcal{L}_!$ be the coKleisli category $\mathcal{L}_!$ of the comonad $!$. Then, by composing the adjunction between \mathcal{L} and $\mathcal{L}_!$ with a state monad $\mathbf{State} \multimap \mathbf{State} \otimes (-)$ on \mathcal{L} , we obtain a monad on $\mathcal{L}_!$:

$$\mathbf{State} \multimap \mathbf{State} \otimes (-) \circlearrowleft \mathcal{L} \xleftarrow{\top} \mathcal{L}_!,$$

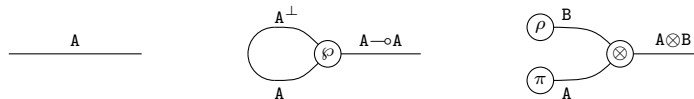
which sends an object $A \in \mathcal{L}_!$ to $\mathbf{State} \multimap \mathbf{State} \otimes !A$. This use of the state monad is motivated by sampling-based operational semantics: we can regard **PCFSS** as a call-by-value λ -calculus with global states consisting of pairs of a non-negative real number and a finite sequence of real numbers, and we can regard **score** and **sample** as effectful operations interacting with those states.

4.1.3 The Third Step

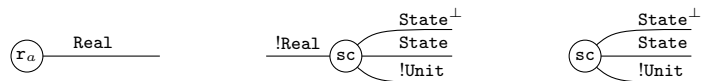
We translate terms in the extended linear λ -calculus into (an extension of proof structures) [37], which are graphical presentations of type derivation trees of linear λ -terms. We can also understand proof structures as string diagrams for compact closed categories [38]. Operators of the pure, linear, λ -calculus, can be translated as usual [37]. For example, type derivation trees

$$\frac{}{x : A \vdash x : A}, \quad \frac{x : A \vdash x : A}{\vdash \lambda x^A. x : A \multimap A} \quad \frac{\pi : \vdash M : A \quad \rho : \vdash N : B}{\vdash M \otimes N : A \otimes B}$$

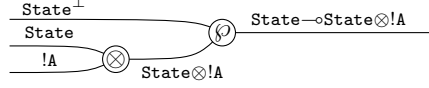
are translated into proof structures



respectively, where nodes labelled with M and N are proof structures associated to type derivations of M and N . Terms of the form r_a , **sample**(M) and **score**, require new kinds of nodes:



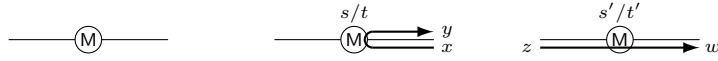
This is not a direct adaptation of typing rules for **score** and **sample** in the linear λ -calculus, but the correspondence can be recovered by way of multiplicatives:



4.2 From Proof Structures to Mealy Machines

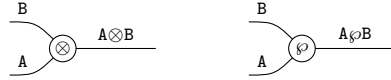
The series of translations from **PCFSS** to proof structures is agnostic as for the computational meaning of **score** and **sample** in **PCFSS** because **score** and **sample** introduced in these translations are just constant symbols. In other words, the translation from **PCFSS** to the extended proof structures is not sound with respect to either form of operational semantics for **PCFSS**. In the last translation step, we assign proof structures a computational meaning, respecting the operational semantics of the underlying **PCFSS** term.

We do this by associating proof structures with Mealy machines. A *Mealy machine* is an input/output-machine whose evolution may depend on its current state. In this paper, for the sake of supporting intuition and of enabling graphical reasoning, we depict a Mealy machine M as a node with some input/output-ports:

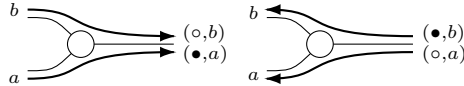


For example, the thick arrow in the middle diagram indicates that if the current state is s and the given input is x , then the Mealy machine outputs y and changes its state to t . In the GoI jargon, data traveling along edges of proof structures are often called *tokens*.

For the standard proof structures, we can follow [39] where Mealy machines associated with proof structures are built up from Mealy machines associated to each nodes. For example, the following nodes

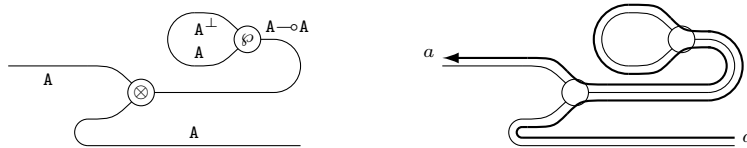


are both associated with a one-state Mealy machine that behaves in the following manner:

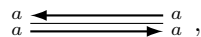


Namely, the Mealy machine forwards each input from the left hand side to the right hand side endowing it with a tag that tells where the token came from. The Mealy machine handles inputs from the right hand side in the reverse way.

Soundness of Mealy machine semantics states that if two (pure) linear λ -terms are β -equivalent, then the behaviours of the Mealy machines associated to these terms are the same. As an example, let us consider a β -reduction step $(\lambda x^A. x) y \rightarrow y$. The proof structure associated to $(\lambda x^A. x) y$ is the graph in the left hand side, and the arrow in the right hand side illustrates a trace of a run of this Mealy machine for an input a from the right edge:

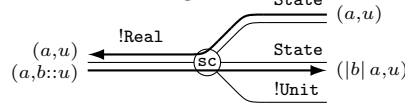


This Mealy machine forwards any input from the right hand side to the left hand side as indicated by the thick arrow, and it also forwards any input from the left hand side to the right hand side. Hence, the behaviour of this Mealy machine is equivalent to the behaviour of the following trivial Mealy machine:



which is the interpretation of $y : A \vdash y : A$. This is in fact a symptom of a general phenomenon: Mealy machine semantics for the linear λ -calculus captures β -reduction $(\lambda x^A. x) y \rightarrow y$.

But how can we extend this Mealy machine semantics to **score** and **sample**? Here, we borrow the idea from Game semantics [40] that models computation in terms of interaction between programs and environments. For scoring and sampling, we can infer how they interact with the environment from sampling-based operational semantics. For scoring, we associate **score** with a one-state Mealy machine that has the following transitions:



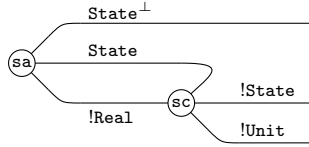
where u is a finite sequence of real numbers and a, b are real numbers such that $a \geq 0$. We can read these transitions as follows: for each “configuration” $(-, a, u)$, the Mealy machine sends a query (a, u) to environment in order to know the value of its argument, and if environment answers that the value is b , i.e., if the Mealy machine receives $(a, b :: u)$, then it outputs $(|b| a, u)$, which is the evaluation result of $(\text{score}(r_b), a, u)$.

For sampling, we associate **sample** with a Mealy machine that has the following transitions:



where u is a finite sequence of real numbers and a, b are real numbers such that $a \geq 0$. The first transition means that in the initial state $*$, given a “configuration” $(-, a, b :: u)$, the Mealy machine pops the first element of $b :: u$ and memorises the value b by changing its state from $*$ to b . After this transition, for any query (a, u) asking the result of sampling, it answers the value memorised in the first transition.

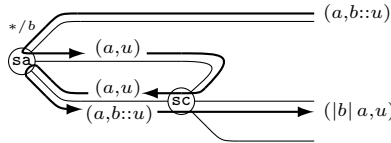
For example, a Mealy machine



which is a denotation of the term

$$M = \text{let } x \text{ be sample in score}(x),$$

and behaves as follows:



Our adequacy theorem says that the evaluation result of a term coincides with the execution result of the associated Mealy machine. In fact, for this case, the outcome $(|b| a, u)$ of the above Mealy machine is equal to the evaluation result of $(M, a, b :: u)$, that is, $(M, a, b :: u) \rightarrow^* (\text{skip}, |b| a, u)$. In this interaction process, the memoisation mechanism of the **sa**-node is necessary, otherwise the **sa**-node can not tell the **sc**-node that the result of sampling is b .

Remark 4.1. Two notions of state (the one coming from the state monad and the one of the Mealy machine itself) are used for different purpose here: the first notion is needed to model the call-by-value evaluation strategy where we need to store intermediate effects that are invoked during the evaluation. The second notion of state is needed to model sampling. More concretely, each Mealy machine for sampling need to remember the already sampled values in the current probabilistic branch.

5 Mealy Machines and their Compositions

After having described Mealy machine semantics briefly and informally, it is now time to get more precise. In this section, we introduce the notion of a Mealy machine and some constructions on Mealy machines. We also introduce a way of diagrammatically presenting Mealy machines which is behaviourally sound.

5.1 Mealy Machines, Formally

In this paper, we call a pair of measurable spaces an **Int-object**. We use sans-serif capital letters X, Y, Z, \dots to denote **Int**-objects, and we denote the positive/negative part of an **Int**-object by the same italic letter superscripted by $+/-$. For example, X denotes an **Int**-object (X^+, X^-) consisting of two measurable spaces X^+ and X^- . The name “**Int**-object” comes from the so-called **Int**-construction [26]. Definition 5.1 and the definition of monoidal products in Section 5.4 are also motivated by **Int**-construction.

Definition 5.1. For **Int**-objects X and Y , a Mealy machine M from X to Y consists of

- a measurable space S_M called the state space of M ;
- an element $s_M \in S_M$ called the initial state of M ;
- a partial measurable function

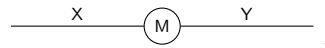
$$\tau_M: (X^+ + Y^-) \times S_M \rightarrow (Y^+ + X^-) \times S_M$$

called the transition function.

If M is a Mealy machine from X to Y , we write $M: X \multimap Y$.

The transition function τ_M of a Mealy machine M describes a mapping between inputs and outputs which can also alter the underlying state. For $x \in X^+ + Y^-$ and $s \in S_M$, $\tau_M(x, s) = (y, t)$ means that when the current state of M is s , given an input x , there is an output y and the next state is t .

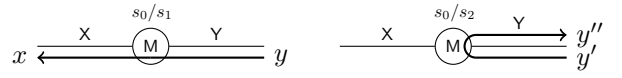
Readers may wonder why X^- appears in the target and Y^- appears in the source of the transition function of a Mealy machine from X to Y . In short, this is because we are interested in Mealy machines that handle bidirectional computation. The diagrammatic presentation of Mealy machines clarifies the meaning of “bidirectional.” Let $M: X \multimap Y$ be a Mealy machine. In this paper, we depict M as follows:



Intuitively, each label on an edge indicates the type of data traveling along the edge. Namely, on the X -edge (on the Y -edge), elements in X^+ (in Y^+) go from left to right, and elements in X^- (in Y^-) go from right to left. For example, we depict the following transitions

$$\tau_M((\circ, y), s_0) = ((\circ, x), s_1), \quad \tau_M((\bullet, y'), s_0) = ((\bullet, y''), s_2)$$

for some $y, y' \in Y^-$, $x \in X^-$, $y'' \in Y^+$ and $s_0, s_1, s_2 \in S_M$ as the following thick arrows



(Recall that the white/black bullet indicates the left/right part of the disjoint sum.) The expressions s_0/s_1 and s_0/s_2 on the Mealy machine M stands for transitions of states. We omit states transitions when we can infer them.

We will give some Mealy machines whose state spaces are trivial, namely 1. We call such a Mealy machine *token machine*. Our usage of the term token machine is along the lines of that

in other papers on GoI such as [41, 39]. Since we can identify the transition function of a token machine $M: X \multimap Y$ with the following partial measurable function

$$X^+ + Y^- \cong (X^+ + Y^-) \times 1 \xrightarrow{\tau_M} (Y^+ + X^-) \times 1 \cong Y^+ + X^-,$$

giving partial measurable function of this type is enough to specify a token machine.

Convention 5.1. *We define a token machine $M: X \multimap Y$ by giving a partial measurable function from $X^+ + Y^-$ to $Y^+ + X^-$, and we also call this partial measurable function transition function of M . Abusing notation, we write τ_M for this transition function.*

5.2 Behavioural Equivalence

We are now ready to give an equivalence relation between Mealy machines which identifies machines which *behave the same way*. Identifying Mealy machines in terms of their behaviour is important to reason about compositions of Mealy machines in the coming sections. Here, we are inspired by behavioural equivalence from coalgebraic theory of modelling transition systems [42].

Let M and N be Mealy machines from X to Y . We write $M \preceq_{X,Y} N$ when there is a measurable function $f: S_M \rightarrow S_N$ satisfying $f(s_M) = s_N$ and

$$\begin{array}{ccc} (X^+ + Y^-) \times S_M & \xrightarrow{\text{id} \times f} & (X^+ + Y^-) \times S_N \\ \tau_M \downarrow & & \downarrow \tau_N \\ (Y^+ + X^-) \times S_M & \xrightarrow{\text{id} \times f} & (Y^+ + X^-) \times S_N \end{array}$$

The definition means that if we have $M \preceq_{X,Y} N$, then no observer can distinguish between M and N from their input/output behaviour, although their internal structure can be quite different. We define an equivalence relation $\simeq_{X,Y}$ to be the reflective, symmetric, and transitive closure of $\preceq_{X,Y}$. Below, if we can infer the subscript X, Y from the context, we write \simeq instead of $\simeq_{X,Y}$.

Definition 5.2. *For Mealy machines $M, N: X \multimap Y$, we say that M is behaviourally equivalent to N when $M \simeq N$.*

For a Mealy machine $M: X \multimap Y$, we write $[M]$ for its equivalence class with respect to behavioural equivalence. We define a binary relation \leq between equivalence classes of Mealy machines from X to Y by $[M] \leq [N]$ if and only if there are $M' \simeq M$ and $N' \simeq N$ such that $S_{M'} = S_{N'}$ and $s_{M'} = s_{N'}$, and the graph relation of $\tau_{M'}$ is a subset of the graph relation of $\tau_{N'}$.

Proposition 5.1. *The set of equivalence classes for $\simeq_{X,Y}$ with \leq is a pointed ω cpo.*

We can characterize interpretation of the fixed point operator in **PCFSS** in terms of least fixed points, see [43]. We give a proof of Proposition 5.1 in Section 5.3.

5.3 Proof of Proposition 5.1

For a partially defined expressions E and E' , we write $E \approx E'$ when E is defined if and only if E' is defined, and if both expressions are defined, then they are the same. For example, we have $(1-x)^{-1} \approx \sum_{n=0}^{\infty} x^n$ for all $x \in \mathbb{R}_{[0,1]}$. For a measurable space X , we write LX for the measurable space of all finite sequences over X endowed with the following σ -algebra:

$$A \in \Sigma_{LX} \iff \text{for all } n \in \mathbb{N}, A \cap X^n \in \Sigma_{X^n}.$$

We write ε for the empty sequence. For $a \in X$ and $u \in LX$, we denote the list obtained by appending a to u by $a :: u$.

Let $M: X \multimap Y$ be a Mealy machine. We write Z for $X^+ + Y^-$ and W for $Y^+ + X^-$. Then, the transition function of M is of the form

$$\tau_M: Z \times S_M \rightarrow W \times S_M.$$

We define partial measurable functions $\alpha_M: LZ \rightarrow S_M$ and $\beta_M: Z \times LZ \rightarrow W$ by

$$\begin{aligned}\alpha_M(\varepsilon) &= s_M, \\ \alpha_M(z :: u) &= \begin{cases} s, & \text{if } \alpha_M(u) \text{ and } \tau_M(z, \alpha_M(u)) \text{ are defined and} \\ & \tau_M(z, \alpha_M(u)) = (w, s) \text{ for some } w \in W, \\ \text{undefined,} & \text{otherwise,} \end{cases} \\ \beta_M(z, u) &= \begin{cases} w, & \text{if } \alpha_M(u) \text{ and } \tau_M(z, \alpha_M(u)) \text{ are defined and} \\ & \tau_M(z, \alpha_M(u)) = (w, s) \text{ for some } s \in S_M, \\ \text{undefined,} & \text{otherwise.} \end{cases}\end{aligned}$$

Below, for $x \in W \times S_M$, we write $\text{fst}(x)$ for the first entry of x , and we write $\text{snd}(x)$ for the second entry of x . By the definition of α_M and β_M , we have

$$\alpha_M(z :: u) \approx \text{snd}(\tau_M(z, \alpha_M(u))), \quad \beta_M(z, u) \approx \text{fst}(\tau_M(z, \alpha_M(u))).$$

Lemma 5.1. *If $M \preceq N$, then $\beta_M = \beta_N$.*

Proof. Let $h: S_M \rightarrow S_N$ be a measurable function that realizes $M \preceq N$. We show $h(\alpha_M(u)) \approx \alpha_N(u)$ and $\beta_M(z, u) \approx \beta_N(z, u)$ by induction on the size of u . (Base case)

$$\begin{aligned}h(\alpha_M(\varepsilon)) &= h(s_M) \\ &= s_N \\ &= \alpha_N(\varepsilon)\end{aligned}$$

$$\begin{aligned}\beta_M(z, \varepsilon) &\approx \text{fst}(\tau_M(z, \alpha_M(\varepsilon))) \\ &\approx \text{fst}((W \times h)(\tau_M(z, \alpha_M(\varepsilon)))) \\ &\approx \text{fst}(\tau_N(z, h(\alpha_M(\varepsilon)))) \\ &\approx \text{fst}(\tau_N(z, \alpha_N(\varepsilon))) \\ &\approx \beta_N(z, \varepsilon).\end{aligned}$$

(Induction step)

$$\begin{aligned}h(\alpha_M(z :: u)) &\approx h(\text{snd}(\tau_M(z, \alpha_M(u)))) \\ &\approx \text{snd}((W \times h)(\tau_M(z, \alpha_M(u)))) \\ &\approx \text{snd}(\tau_N(z, h(\alpha_M(u)))) \\ &\approx \text{snd}(\tau_N(z, \alpha_N(u))) \\ &\approx \alpha_N(z :: u).\end{aligned}$$

$$\begin{aligned}\beta_M(z', z :: u) &\approx \text{fst}(\tau_M(z', \alpha_M(z :: u))) \\ &\approx \text{fst}((W \times h)(\tau_M(z', \alpha_M(z :: u)))) \\ &\approx \text{fst}(\tau_N(z', h(\alpha_M(z :: u)))) \\ &\approx \text{fst}(\tau_N(z', \alpha_N(z :: u))) \\ &\approx \beta_N(z', z :: u).\end{aligned}$$

□

For a Mealy machine $M: X \multimap Y$, we define Mealy machines $M^\#$, $M^\circ: X \multimap Y$ by

- $S_{M^\#} = LZ$,
- $s_{M^\#} = \varepsilon$,

- $\tau_{M^\#}(z, u) = \begin{cases} (\beta_M(z, u), z :: u), & \text{if } \beta_M(z, u) \text{ is defined,} \\ \text{undefined,} & \text{otherwise,} \end{cases}$

and

- $S_{M^\oplus} = \{\diamond\} \cup S_M,$
- $s_{M^\oplus} = s_M,$
- $\tau_{M^\oplus}(z, s) = \begin{cases} \tau_M(z, s), & \text{if } s \in S_M \text{ and } \tau_M(z, s) \text{ is defined,} \\ \text{undefined,} & \text{if } u = \diamond, \\ \text{undefined,} & \text{otherwise.} \end{cases}$

Here, the σ -algebra of S_{M^\oplus} is the one induced by Σ_{1+S_M} via the obvious bijection between $1 + S_M$ and $\{\diamond\} \cup S_M$.

Lemma 5.2. $M \preceq M^\oplus \succeq M^\#$.

Proof. It is straightforward to check that the embedding $e: S_M \rightarrow \{\diamond\} \cup S_M$ is a measurable function that realizes $M \preceq M^\oplus$. It remains to show $M^\# \preceq M^\oplus$. We define a measurable function $h: LZ \rightarrow \{\diamond\} \cup S_M$ by

$$h(u) = \begin{cases} \alpha_M(u), & \text{if } \alpha_M(u) \text{ is defined,} \\ \diamond, & \text{otherwise.} \end{cases}$$

We show that for any $(z, u) \in Z \times LZ$,

$$\tau_{M^\oplus}((Z \times h)(z, u)) \approx (W \times h)(\tau_{M^\#}(z, u))$$

by induction on $u \in LZ$. (Base case)

$$\begin{aligned} \tau_{M^\oplus}(z, h(\varepsilon)) &= \tau_{M^\oplus}(z, s_M) = \tau_M(z, s_M) \approx (\text{fst}(\tau_M(z, \alpha_M(\varepsilon))), \text{snd}(\tau_M(z, \alpha_M(\varepsilon)))) \\ &\approx (\text{fst}(\tau_M(z, \alpha_M(\varepsilon))), \alpha_M(z :: \varepsilon)) \\ &\approx (\beta_M(z, \varepsilon), h(z :: \varepsilon)) \\ &\approx (W \times h)(\tau_{M^\#}(z, \varepsilon)). \end{aligned}$$

(Induction step)

$$\begin{aligned} \tau_{M^\oplus}((Z \times h)(z, z' :: u)) &\approx \tau_M(z, \alpha_M(z' :: u)) \\ &\approx \tau_M(z, \text{snd}(\tau_M(z', \alpha_M(u)))) \\ &\approx \tau_{M^\oplus}(z, \text{snd}(\tau_M(z', \alpha_M(u)))) \\ &\approx \tau_{M^\oplus}(z, \text{snd}(\tau_{M^\oplus}(z', h(u)))) \\ &\approx \tau_{M^\oplus}(z, h(\text{snd}(\tau_{M^\oplus}(z', u)))) \\ &\approx \tau_M(z, \alpha_M(\text{snd}(\tau_{M^\oplus}(z', u)))) \\ &\approx \tau_M(z, \alpha_M(z' :: u)) \\ &\approx (\beta_M(z, z' :: u), \alpha_M(z :: z' :: u)) \\ &\approx (\beta_M(z, z' :: u), h(z :: z' :: u)) \\ &\approx (W \times h)(\tau_{M^\#}(z, z' :: u)). \end{aligned}$$

□

Proposition 5.2. For all Mealy machines $M, N: X \multimap Y$, we have $M \simeq N$ if and only if $\beta_M = \beta_N$.

Proof. If $M \simeq N$, then we can show that $\beta_M = \beta_N$ by using Lemma 5.1. If $\beta_M = \beta_N$, then we have $M^\# = N^\#$ by the definition of $(-)^{\#}$. Because we have $M \simeq M^\#$ and $N \simeq N^\#$ (Lemma 5.2), we see that M is behaviourally equivalent to N . □

Hence, each equivalence class $[M]$ of behavioural equivalence is represented by $M^\#$, and $M^\#$ is independent of choice of M . We extend this correspondence to order theoretic structure of Mealy machines.

Lemma 5.3. *Let M, N be Mealy machines from X to Y such that $S_M = S_N$. If $\tau_M \leq \tau_N$ and $s_M = s_N$, then $\tau_{M^\#} \leq \tau_{N^\#}$.*

Proof. By induction on the size of $u \in LZ$, we can show that if $\alpha_M(u)$ is defined, then $\alpha_N(u)$ is defined and they are the same. Then $\tau_{M^\#} \leq \tau_{N^\#}$ follows from the definition of $(-)^{\#}$. \square

Theorem 5.1. *For Mealy machines $M, N: X \multimap Y$,*

$$[M] \leq [N] \iff \tau_{M^\#} \leq \tau_{N^\#}.$$

Proof. If $\tau_{M^\#} \leq \tau_{N^\#}$, then because $M^\#$ and $N^\#$ are representatives of $[M]$ and $[N]$ respectively, we have $[M] \leq [N]$. If $[M] \leq [N]$, then there are $M' \simeq M$ and $N' \simeq N$ such that

- $S_{M'} = S_{N'}$ and $s_{M'} = s_{N'}$,
- the graph relation of $\tau_{M'}$ is a subset of the graph relation of $\tau_{N'}$.

By Lemma 5.3, we see that $\tau_{M'^\#} \leq \tau_{N'^\#}$. \square

Theorem 5.2. *The set of equivalence classes of Mealy machines from X to Y with the partial order \leq is an ω -cpo.*

Proof. Let $[N]$ be an upper bound of an ω -chain

$$[M_1] \leq [M_2] \leq \dots$$

By Theorem 5.1, we have

$$\tau_{M_1^\#} \leq \tau_{M_2^\#} \leq \dots \leq \tau_{N^\#}.$$

We define a Mealy machine $L: X \multimap Y$ by

- $S_L = S_{M_1^\#}$,
- $s_L = s_{M_1^\#}$,
- $\tau_L = \bigvee_{n \in \mathbb{N}} \tau_{M_n^\#}$.

Because $M_n \simeq M_n^\#$, the equivalence class $[L]$ is an upper bound of the ω -chain $[M_1] \leq [M_2] \leq \dots$. We also have $[L] \leq [N]$ because $\tau_L \leq \tau_{M_n^\#}$. \square

5.4 Constructions on Mealy Machines

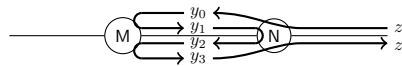
It is now time to give some constructions which are the basic building blocks of our Mealy machine semantics. This section consists of three parts. The first part (from Section 5.4.2 to Section 5.4.5) is related to the linear λ -calculus and serves to model the purely functional features of **PCFSS**, such as λ -abstraction and function application. In the second part (Section 5.4.6 and Section 5.4.7), we give Mealy machines modelling real numbers and measurable functions. In the last part (from Section 5.4.9 to Section 5.4.11), we introduce a state monad and associate the monad with Mealy machines modelling **score** and **sample**.

5.4.1 Composition

Let X, Y and Z be **Int**-objects, and let $M: X \multimap Y, N: Y \multimap Z$ be Mealy machines. We can now define their composition $N \circ M: X \multimap Z$. Before giving a precise definition, some intuitive explanation about $N \circ M$ is in order. The main idea is to define $N \circ M$ as a Mealy machine obtained by connecting N and M in the following manner:



The following series of thick arrows



illustrates an execution of the obtained Mealy machine. Given an input from an edge, M and N engage in some interactive communication, and at some point, some output is produced. Because $N \circ M$ performs “parallel composition plus connecting”, the state space of $N \circ M$ should be $S_M \times S_N$, and the initial state should be (s_M, s_N) . The transition function of $N \circ M$ should be given by the collection of all possible interaction paths between M and N .

Let us give a precise definition. For Mealy machines $M: X \multimap Y$ and $N: Y \multimap Z$, we define the state space and the initial states of $N \circ M$ by $S_{N \circ M} = S_M \times S_N$, $s_{N \circ M} = (s_M, s_N)$ and we define the transition function $\tau_{N \circ M}$ by

$$\tau_{N \circ M} = f_{X^+, Z^-, Z^+, X^-} \vee \bigvee_{n \in \mathbb{N}} f_{Y^+, Y^-, Z^+, X^-} \circ f_{Y^+, Y^-, Y^+, Y^-}^n \circ f_{X^+, Z^-, Y^+, Y^-}$$

where the $f_{A,B,C,D}: (A + B) \times S_{N \circ M} \rightarrow (C + D) \times S_{N \circ M}$ are restrictions of the following partial measurable function

$$\begin{aligned} & (X^+ + Z^- + Y^+ + Y^-) \times S_{N \circ M} \\ & \quad \downarrow \cong \\ & (X^+ + Y^-) \times S_M \times S_N + (Y^+ + Z^-) \times S_N \times S_M \\ & \quad \downarrow \tau_M \times S_N + \tau_N \times S_M \\ & (Y^+ + X^-) \times S_M \times S_N + (Z^+ + Y^-) \times S_N \times S_M \\ & \quad \downarrow \cong \\ & (Z^+ + X^- + Y^+ + Y^-) \times S_{N \circ M}, \end{aligned}$$

and the above join is with respect to the inclusion order between graph relations. The above join is measurable because measurable sets are closed under countable joins. It is tedious but doable to check that the above join always exists and that the composition is compatible with behavioural equivalence and satisfies associativity modulo behavioural equivalence. We define a token machine $\text{id}_X: X \multimap X$ by $\tau_{\text{id}_X} = \text{id}_{X^+ + X^-}$. This is the unit of the composition modulo behavioural equivalence.

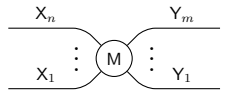
5.4.2 Monoidal Products

Monoidal Products of Int-objects We introduce monoidal products of **Int**-objects and their diagrammatic presentation. For **Int**-objects X and Y , we define an **Int**-object $X \otimes Y$ by

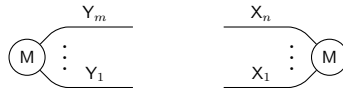
$$X \otimes Y = (X^+ + Y^+, Y^- + X^-).$$

We define an **Int**-object I to be (\emptyset, \emptyset) . We write $X \otimes Y \otimes \dots$ for $X \otimes (Y \otimes \dots)$.

Let $X_1, \dots, X_n, Y_1, \dots, Y_m$ be **Int**-object. We depict a Mealy machine M from $X_1 \otimes \dots \otimes X_n$ to $Y_1 \otimes \dots \otimes Y_m$ as a node with edges labeled by X_1, \dots, X_n on the left hand side and edges labeled by Y_1, \dots, Y_m on the right hand side:



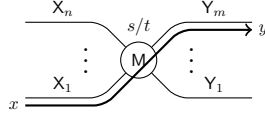
We do not draw any edges on the left/right hand side when the domain/codomain of M is I :



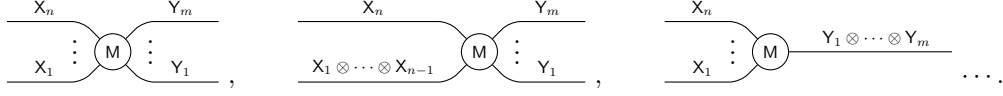
The diagrammatic presentation of monoidal products allows for an intuitive description of transition functions. For example, we can depict a transition

$$\tau_M((\bullet, (\bullet, x)), s) = ((\bullet, (\circ, \dots (\circ, y))), t)$$

for some $x \in X_1^+$, $y \in Y_m^+$ and $s, t \in S_M$ as follows:



We note that there are several ways to present a Mealy machine $M: X_1 \otimes \dots \otimes X_n \multimap Y_1 \otimes \dots \otimes Y_m$ such as

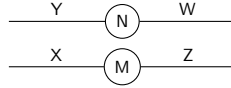


Monoidal Product of Mealy Machines We give monoidal products of Mealy machines. For Mealy machines $M: X \multimap Z$ and $N: Y \multimap W$, we define a Mealy machine $M \otimes N: X \otimes Y \multimap Z \otimes W$ by: $S_{M \otimes N} = S_M \times S_N$, $s_{M \otimes N} = (s_M, s_N)$ and $\tau_{M \otimes N}$ is given by

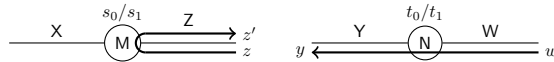
$$\begin{aligned}
 & ((X^+ + Y^+) + (W^- + Z^-)) \times S_{M \otimes N} \\
 & \quad \downarrow \cong \\
 & (X^+ + Z^-) \times S_M \times S_N + (Y^+ + W^-) \times S_N \times S_M \\
 & \quad \downarrow \tau_M \times S_N + \tau_N \times S_M \\
 & (Z^+ + X^-) \times S_M \times S_N + (W^+ + Y^-) \times S_N \times S_M \\
 & \quad \downarrow \cong \\
 & ((Z^+ + W^+) + (Y^- + X^-)) \times S_{M \otimes N} .
 \end{aligned}$$

It is not difficult to check that the monoidal product is compatible with behavioural equivalence.

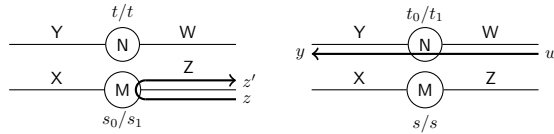
We depict $M \otimes N: (X \otimes Y) \multimap (Z \otimes W)$ as follows:



As indicated by the above diagram, $M \otimes N$ consists of two sub-machines M and N working independently. For example, if we have



then $M \otimes N$ has the following transitions:



for all $t \in S_N$ and for all $s \in S_M$.

Convention 5.2. We do the following identification:

- We identify $X \otimes (Y \otimes Z)$ with $(X \otimes Y) \otimes Z$ by the canonical isomorphism $X + (Y + Z) \cong (X + Y) + Z$.
- We identify $I \otimes X$ and $X \otimes I$ with X by the unit laws $X^+ + \emptyset \cong X^+$ and $\emptyset + X^- \cong X^-$.

5.4.3 Axiom Link and Cut Link

For an **Int**-object X , we define X^\perp to be (X^-, X^+) , and we define token machines

$$\text{unit}_X: I \multimap X \otimes X^\perp, \quad \text{counit}_X: X^\perp \otimes X \multimap I$$

by $\tau_{\text{unit}_X} = \text{id}_{X^+ + X^-}$ and $\tau_{\text{counit}_X} = \text{id}_{X^- + X^+}$. We depict them by single edges



respectively. This is compatible with behaviour of these Mealy machines: if we give an input to an edge, then we will get the same value from the other end of the edge. For example, for any $x \in X^+$, we have



5.4.4 Symmetry

Let X and Y be **Int**-objects. We define a token machine $\text{sym}_{X,Y}: X \otimes Y \multimap Y \otimes X$ by letting its transition function be the canonical isomorphism

$$(X^+ + Y^+) + (X^- + Y^-) \xrightarrow{\cong} (Y^+ + X^+) + (Y^- + X^-).$$

We depict $\text{sym}_{X,Y}$ by a crossing:



As arrows in the right hand side indicate, given an input from an edge in one side, $\text{sym}_{X,Y}$ outputs the same value to the corresponding edge on other side.

5.4.5 A Modal Operator

We give a constructor on Mealy machines that corresponds to the resource modality in linear logic. For an **Int**-object X , we define an **Int**-object $!X$ by

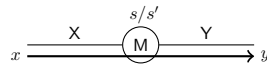
$$!X = (\mathbb{N} \times X^+, \mathbb{N} \times X^-).$$

We can informally regard $!X$ as a countable monoidal power $\bigotimes_{n \in \mathbb{N}} X \approx X \otimes X \otimes \dots$. Following this intuition, we extend the action of $!(-)$ to Mealy machines. Let $M: X \multimap Y$ be a Mealy machine. We define a Mealy machine $!M: !X \multimap !Y$ by: the state space of $!M$ is defined to be $|\mathbb{M}|^{\mathbb{N}}$ associated with the least σ -algebra such that for all $A_1, A_2, \dots \in \Sigma_M$, $A_1 \times A_2 \times \dots \in \Sigma_{!M}$; the initial state $s_{!M}$ is (s_M, s_M, \dots) ; the transition function $\tau_{!M}$ is the unique partial measurable function satisfying

$$\begin{array}{ccc} (X^+ + Y^-) \times S_M \times S_M^{\mathbb{N}} & \xrightarrow{(\text{inj}_n + \text{inj}_n) \times \text{ins}_n} & (\mathbb{N} \times X^+ + \mathbb{N} \times Y^-) \times S_M^{\mathbb{N}} \\ \tau_M \times S_M^{\mathbb{N}} \downarrow & & \downarrow \tau_{!M} \\ (Y^+ + X^-) \times S_M \times S_M^{\mathbb{N}} & \xrightarrow{(\text{inj}_n + \text{inj}_n) \times \text{ins}_n} & (\mathbb{N} \times Y^+ + \mathbb{N} \times X^-) \times S_M^{\mathbb{N}} \end{array}$$

for all $n \in \mathbb{N}$. Here, $\text{inj}_n: (-) \rightarrow \mathbb{N} \times (-)$ are the n th injections, and $\text{ins}_n: S_M \times S_M^{\mathbb{N}} \rightarrow S_M^{\mathbb{N}}$ sends $(s, \{s_n\}_{n \in \mathbb{N}})$ to $(s_0, \dots, s_{n-1}, s, s_n, s_{n+1}, \dots)$.

As $!(-)$ is defined to be a countable monoidal power, $!M$ behaves as a parallel composition of countably infinite copies of M . For example, if we have



then for all $n \in \mathbb{N}$ and $t_1, t_2, \dots \in S_M$, we have

$$\begin{array}{ccc} (t_1, \dots, t_{n-1}, s, t_n, t_{n+1}, \dots) & / & (t_1, \dots, t_{n-1}, s', t_n, t_{n+1}, \dots) \\ (n, x) & \xrightarrow{!M} & (n, y) \end{array}$$

In other words, given an input whose first entry is n , then the n th copy of M handles the input, and there is no side effect to the other copies of M .

Proposition 5.3. *The operator $!(-)$ is compatible with the behavioral equivalence and is functorial. Namely,*

- for all Mealy machines $M, N: X \multimap Y$, if $M \simeq M'$, then $!M \simeq !N$; and
- for all Mealy machines $M: X \multimap Y$ and $N: Y \multimap Z$,

$$!(N \circ M) \simeq !N \circ !M;$$

- $!id_X \simeq id_{!X}$.

Convention 5.3. *For the sake of legibility and due to lack of space, we sometimes implicitly identify $!(X \otimes Y)$ with $!X \otimes !Y$ by the canonical isomorphism $\mathbb{N} \times (X + Y) \cong \mathbb{N} \times X + \mathbb{N} \times Y$.*

Under the above convention, for Mealy machines $M: !(X \otimes Y) \multimap Z$ and $N: W \multimap X$, we can simply write $M \circ (!N \otimes Y_{!Y}): !W \otimes !Y \multimap Z$. It is not difficult to see that when $Z = !Z'$ and $M = !M'$ for some $M': X \otimes Y \multimap Z'$, we have $M \circ (!N \otimes id_{!Y}) \simeq !(M \circ (N \otimes id_Y))$.

Dereliction For an **Int**-object X , we define a token machine $d_X: !X \multimap X$ by defining $\tau_{d_X}: (\mathbb{N} \times X^+) + X^- \rightarrow X^+ + (\mathbb{N} \times X^-)$ by

$$\tau_{d_X}(\bullet, (n, x)) = (\bullet, x), \quad \tau_{d_X}(\circ, x) = (\circ, (0, x)).$$

The Mealy machine d_X pops/pushes indices with probability 1. Namely, we have

$$(n, x) \xrightarrow{!X} \textcircled{d_X} \xrightarrow{X} x \quad (0, x) \xleftarrow{!X} \textcircled{d_X} \xleftarrow{X} x$$

for all $n \in \mathbb{N}$, $x \in X^+$ and $x' \in X^-$. Hence, for any Mealy machine $M: I \multimap X$, if we have

$$\textcircled{M} \xrightarrow{s/s'} \textcircled{X} \xrightarrow{x} x'$$

for some $x \in X^-$, $x' \in X^+$ and $s, s' \in S_M$, then $d_X \circ !M$ has the following transition:

$$(s, s_1, s_2, \dots) / (s', s_1, s_2, \dots) \textcircled{!M} \xrightarrow{!X} \textcircled{d_X} \xrightarrow{X} x' \quad \textcircled{d_X} \xleftarrow{X} x \xleftarrow{!X} (0, x)$$

for all $s_1, s_2, \dots \in S_M$.

Proposition 5.4. *For any Mealy machine $M: I \multimap X$,*

$$d_X \circ !M \simeq M.$$

Diagrammatically, we have

$$\textcircled{!M} \xrightarrow{!X} \textcircled{d_X} \xrightarrow{X} \simeq \textcircled{M} \xrightarrow{X}.$$

Digging and Contraction For natural numbers $n, m \in \mathbb{N}$, we write $\langle n, m \rangle$ for the Cantor pairing $n + (n + m)(n + m + 1)/2$, and we write $n|_0$ and $n|_1$ for unique natural numbers such that $n = \langle n|_0, n|_1 \rangle$. For an **Int**-object X , let $dg_X: !X \multimap !X$ and $con_X: !X \multimap !X \otimes !X$ be stateless deterministic Mealy machines whose transition functions

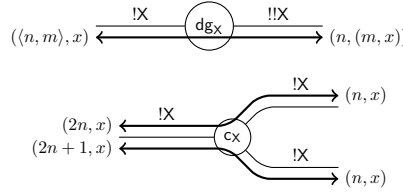
$$\tau_{dg_X}: \mathbb{N} \times X^+ + \mathbb{N} \times \mathbb{N} \times X^- \rightarrow \mathbb{N} \times \mathbb{N} \times X^+ + \mathbb{N} \times X^-,$$

$$\tau_{con_X}: \mathbb{N} \times X^+ + (\mathbb{N} \times X^- + \mathbb{N} \times X^-) \rightarrow (\mathbb{N} \times X^+ + \mathbb{N} \times X^+) + \mathbb{N} \times X^-.$$

are given by

$$\begin{aligned}
\tau_{\text{dg}_X}(\bullet, (\langle n, m \rangle, x)) &= (\bullet, (n, m, x)) \\
\tau_{\text{dg}_X}(\circ, (n, m, x)) &= (\circ, (\langle n, m \rangle, x)), \\
\tau_{\text{con}_X}(\bullet, (n, x)) &= \begin{cases} (\bullet, (\bullet, (n/2, x))), & \text{if } n \text{ is even,} \\ (\bullet, (1, ((n-1)/2, x))), & \text{if } n \text{ is odd,} \end{cases} \\
\tau_{\text{con}_X}(\circ, (\bullet, (u, x))) &= (\circ, (2n+1, x)), \\
\tau_{\text{con}_X}(\circ, (\circ, (u, x))) &= (\circ, (2n, x)).
\end{aligned}$$

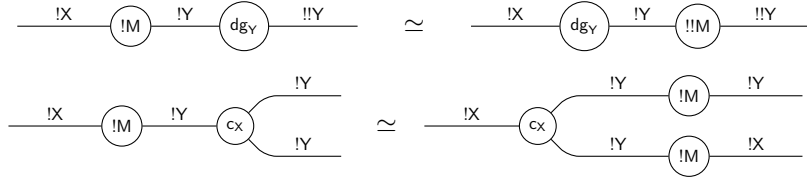
These stateless Mealy machines dg_X and con_X behave as follows: for all $n, m \in \mathbb{N}$,



Proposition 5.5. For any Mealy machine $M: X \multimap Y$,

$$\text{dg}_Y \circ !M \simeq !!M \circ \text{dg}_X, \quad \text{con}_Y \circ !M \simeq (!M \otimes !M) \circ \text{con}_X.$$

Diagrammatically, we have



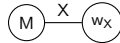
Weakening We define a token machine $w_X: X \rightarrow \mathbf{I}$ by

$$\tau_{w_X} = \text{the empty partial function.}$$

Because the identity is the only Mealy machine from \mathbf{I} to \mathbf{I} (up to behavioural equivalence), we see that for any Mealy machine $M: \mathbf{I} \multimap X$,

$$w_X \circ M \simeq \text{id}_{\mathbf{I}}.$$

This behavioural equivalence means that we can remove



from any diagram.

5.4.6 Real Numbers

We define an **Int**-object R to be (\mathbb{S}, \mathbb{S}) where \mathbb{S} is the measurable space of all finite sequences of real numbers endowed with the following σ -algebra

$$A \in \Sigma_{\mathbb{S}} \iff A \cap \mathbb{R}^n \in \Sigma_{\mathbb{R}^n} \text{ for all } n \in \mathbb{N}.$$

For $a \in \mathbb{R}$, we define a token machine $r_a: \mathbf{I} \multimap R$ by

$$\tau_{r_a}(\circ, u) = (\bullet, a :: u) \quad \text{with a diagram showing } r_a \text{ as a circle with an arrow labeled } R \text{ pointing to } a :: u \text{ and an arrow labeled } u \text{ pointing to } r_a.$$

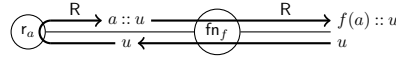
The transition means that given a query u from environment, r_a answers its value a by appending a to u . We will use u as a stack. See Section 5.4.7 and Section 5.4.10.

5.4.7 Measurable Functions

We associate a measurable function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ with a token machine $\text{fn}_f: \mathbb{R}^{\otimes n} \multimap \mathbb{R}$. For simplicity, we define fn_f for $n = 1$ and $n = 2$. When $n = 1$, the transition function $\tau_{\text{fn}_f}: \mathbb{S} + \mathbb{S} \rightarrow \mathbb{S} + \mathbb{S}$ is given by

$$\begin{aligned} \tau_{\text{fn}_f}(\circ, u) &= (\circ, u), \\ \tau_{\text{fn}_f}(\bullet, u) &= \begin{cases} (\bullet, f(a) :: u'), & \text{if } u = a :: u', \\ \text{undefined}, & \text{otherwise.} \end{cases} \end{aligned}$$

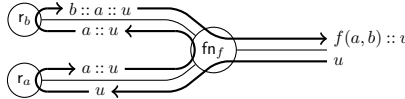
We explain how fn_f simulates f by describing execution of $\text{fn}_f \circ r_a$ for a real number $a \in \mathbb{R}$. As in the following diagram, given an input $u \in \mathbb{S}$ from the right R-edge, fn_f sends u to the left R-edge in order to obtain the value of its argument. The return value to fn_f from r_a is $a :: u$, by which fn_f sees that its argument is a . Then, fn_f outputs $f(a) :: u$. As a whole, the following Mealy machine is behaviourally equivalent to $r_{f(a)}$.



When $n = 2$, the transition function of $\text{fn}_f: \mathbb{R} \otimes \mathbb{R} \multimap \mathbb{R}$ is $\tau_{\text{fn}_f}: (\mathbb{S} + \mathbb{S}) + \mathbb{S} \rightarrow \mathbb{S} + (\mathbb{S} + \mathbb{S})$ given by

$$\begin{aligned} \tau_{\text{fn}_f}(\bullet, (\bullet, u)) &= (\circ, (\bullet, u)), \\ \tau_{\text{fn}_f}(\bullet, (\circ, u)) &= \begin{cases} (\bullet, f(a, b) :: v), & \text{if } u = a :: b :: v, \\ \text{undefined}, & \text{otherwise,} \end{cases} \\ \tau_{\text{fn}_f}(\circ, u) &= (\circ, (\circ, u)). \end{aligned}$$

As in the following diagram, given an input $u \in \mathbb{S}$ from the right R-edge, fn_f first sends u to the lower R-edge in the left hand side in order to obtain the value of its first argument. The return value to fn_f from r_a is $a :: u$. Next, fn_f sends $a :: u$ to the upper R-edge in the left hand side. Then r_b returns $b :: a :: u$. Now, fn_f sees that its first argument is a and its second argument is b . Finally, fn_f outputs $f(a, b) :: u$.



For general cases, f may have more arguments, and fn_f sequentially sends queries to its arguments storing partial information about its arguments on finite sequences of real numbers.

5.4.8 Conditional Branching

For an **Int**-object X such that X^- is a measurable subspace of \mathbb{S} , we define

$$\text{cd}: \mathbb{R} \otimes (X \otimes X) \rightarrow X$$

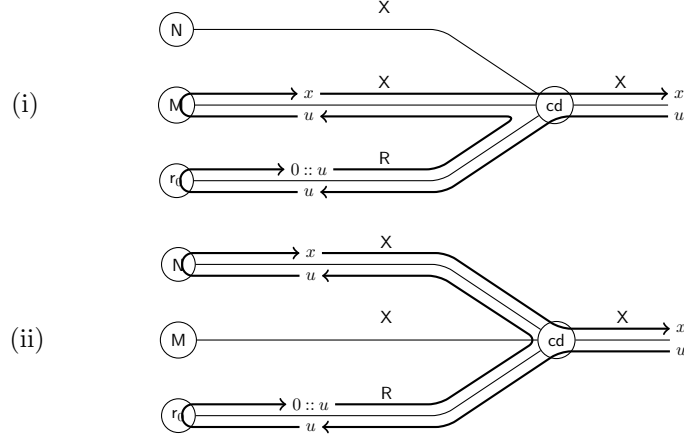
to be a token machine whose transition function

$$\tau_{\text{cd}_X}: (\mathbb{S} + (X^+ + X^+)) + X^- \rightarrow X^+ + ((X^- + X^-) + \mathbb{S})$$

is given by

$$\begin{aligned} \tau_{\text{cd}_X}(\bullet, (\bullet, u)) &= \begin{cases} (\circ, (\bullet, (\circ, v))), & \text{if } u = 0 :: v \text{ and } v \in X^-, \\ (\circ, (\bullet, (\bullet, v))), & \text{if } u = a :: v \text{ and } a \neq 0 \text{ and } v \in X^-, \\ \text{undefined}, & \text{otherwise,} \end{cases} \\ \tau_{\text{cd}_X}(\bullet, (\circ, (\bullet, x))) &= (\bullet, x), \\ (\bullet, (\circ, (\circ, x))) &= (\bullet, x), \\ \tau_{\text{cd}_X}(\circ, u) &= (\circ, (\circ, u)). \end{aligned}$$

For a real number $a \in \mathbb{R}$ and Mealy machines $M, N: I \multimap X$, we describe execution of $\text{cd}_X \circ (r_a \otimes M \otimes N)$. Given an input $u \in X^-$, then cd_X tries to check whether a is zero or not by sending u to the R-edge. There are two cases: (i) if a is 0, then r_a returns $0 :: u$, and cd_X forwards u to the middle X-edge; (ii) if a is not 0, say 1, then r_a returns $1 :: u$, and cd_X forwards u to the upper X-edge:

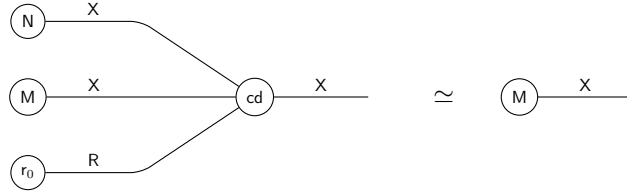


Because in both cases, all outputs from M and N are sent to the X-edge in the right hand, we see that $\text{cd}_X \circ (r_a \otimes M \otimes N)$ simulates M when $a = 0$ and simulates N when $a \neq 0$.

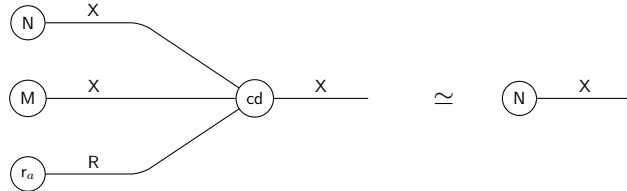
Proposition 5.6. *For $a \in \mathbb{R}$ and for Mealy machines $M, N: I \multimap X$, we have*

$$\text{cd}_X \circ (r_a \otimes M \otimes N) \simeq \begin{cases} M, & \text{if } a = 0, \\ N, & \text{if } a \neq 0. \end{cases}$$

Diagrammatically, we have



and for any $a \neq 0$,



Proof. When $a = 0$, the first behavioral equivalence is realized by the first projection from $1 \times 1 \times S_M \times S_M \cong S_M \times S_M$ to S_M . When $a \neq 0$, the first behavioral equivalence is realized by the second projection from $1 \times 1 \times S_M \times S_M \cong S_M \times S_M$ to S_N . The second behavioral equivalence is realized by the obvious measurable function from $1 \times 1 \times S_M \times S_N$ to 1. \square

5.4.9 A State Monad

Let \mathbb{T} be the subspace of \mathbb{S} consisting of all finite sequences of real numbers in $\mathbb{R}_{[0,1]}$. Recall that $\mathbb{R}_{\geq 0} \times \mathbb{T}$ is “the set of states” in sampling-based operational semantics and our idea is to model

score and **sample** by a state monad. In this section, we give a state monad that we use in our Mealy machine semantics. We define **Int**-objects S_0 and S by

$$S_0 = (\mathbb{R}_{\geq 0} \times \mathbb{T}, \emptyset), \quad S = (\mathbb{R}_{\geq 0} \times \mathbb{T}, \mathbb{R}_{\geq 0} \times \mathbb{T}).$$

Then $S \otimes (-)$ is a state monad (on **Mealy**) because for any **Int**-object X , we have $S \otimes X = ((S_0 \otimes X)^\perp \otimes S_0)^\perp$. The unit and the multiplication of this monad are:

$$e \otimes X: X \multimap S \otimes X, \quad m \otimes X: S \otimes S \otimes X \multimap S \otimes X$$

where $e = \text{units}_{S_0}$ and $m = S_0 \otimes \text{counits}_{S_0} \otimes S_0^\perp$. Note that S is equal to $S_0 \otimes S_0^\perp$. We can depict the unit and the multiplication as follows:

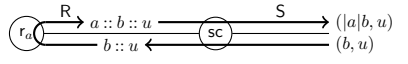


5.4.10 Scoring

We define **sc** to be a token machine from R to S whose transition function $\tau_{\text{sc}}: \mathbb{S} + \mathbb{R}_{\geq 0} \times \mathbb{T} \rightarrow \mathbb{R}_{\geq 0} \times \mathbb{T} + \mathbb{S}$ is given by

$$\begin{aligned} \tau_{\text{sc}}(\circ, (a, u)) &= (\circ, a :: u), \\ \tau_{\text{sc}}(\bullet, u) &= \begin{cases} (\bullet, (|ab|, u')), & \text{if } u = a :: b :: u' \text{ and } u' \in \mathbb{T}, \\ \text{undefined}, & \text{otherwise.} \end{cases} \end{aligned}$$

The token machine simulates scoring ($\text{score}(r_a), b, u \rightarrow (\text{skip}, |a| b, u)$) as follows:



5.4.11 Sampling

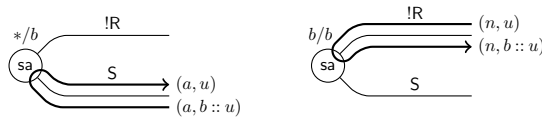
We define a Mealy machine **sa**: $I \multimap S \otimes !R$ by: the state space S_{sa} is defined to be $\{*\} \cup \mathbb{R}_{[0,1]}$, and the initial state s_{sa} is $*$, and the transition function

$$\tau_{\text{sa}}: (\emptyset + (\mathbb{N} \times \mathbb{S} + \mathbb{R}_{\geq 0} \times \mathbb{T})) \times S_{\text{sa}} \rightarrow ((\mathbb{R}_{\geq 0} \times \mathbb{T} + \mathbb{N} \times \mathbb{S}) + \emptyset) \times S_{\text{sa}}$$

is given by

$$\begin{aligned} \tau_{\text{sa}}((\circ, (\bullet, (n, u))), s) &= \begin{cases} \text{undefined}, & \text{if } s = *, \\ ((\bullet, (\circ, (n, s :: u))), s), & \text{if } s \in \mathbb{R}, \end{cases} \\ \tau_{\text{sa}}((\circ, (\circ, (a, u))), s) &= \begin{cases} ((\bullet, (\bullet, (a, v))), b), & \text{if } s = * \text{ and } u = b :: v \\ \text{undefined}, & \text{otherwise.} \end{cases} \end{aligned}$$

As we explained in Section 4.2, the Mealy machine **sa** simulates the evaluation rule ($\text{sample}, a, b :: u \rightarrow (b, a, u)$):



Namely, **sa** pops b from the trace, and then **sa** answers queries (n, u) that the result of sampling is b .

5.5 Diagrammatic Reasoning

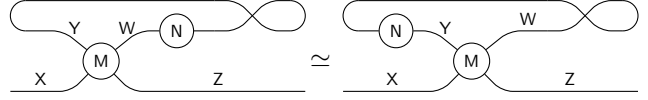
We now give a brief remark on diagrammatic presentation of Mealy machines. The diagrammatic presentation of a Mealy machine is not only for intuitive explanation, but also for rigorous reasoning about behavioural equivalence. This follows from some categorical observation. Let **Mealy** be the category of **Int**-objects and behavioural equivalence classes of Mealy machines, where composition is induced by the composition of Mealy machines. We will give a proof of the following proposition in the next section.

Proposition 5.7. *The category **Mealy** is a compact closed category. The dual of an **Int**-object X is X^\perp . The unit and the counit arrows are unit_X and counit_X .*

Therefore, as a consequence of the coherence theorem for compact closed categories [44, 38], we see that graph isomorphism preserves behavioural equivalence.

Proposition 5.8. *If two Mealy machines have the same diagrammatic presentation modulo some rearrangement of edges and nodes, then they are behaviourally equivalent.*

For example, for all Mealy machines $M: X \otimes Y \multimap Z \otimes W$ and $N: W \multimap Y$, we have



5.6 Proof of Proposition 5.7

5.6.1 The Category of Partial Measurable Functions

For some basic categorical notions, see standard text books such as [45]. We define **pMeas** to be the category of measurable spaces and partial measurable functions. In **pMeas**, the empty space \emptyset is the initial object, and the coproduct space $X + Y$ is the coproduct of X and Y in the categorical sense. We write

$$\text{inl}_{X,Y}: X \rightarrow X + Y, \quad \text{inr}_{X,Y}: Y \rightarrow X + Y$$

for the left/right injections. For partial measurable functions $f: X \rightarrow Y$ and $g: Z \rightarrow Y$, we define $[f, g]: X + Z \rightarrow Y$ to be the cotupling of f and g . For partial measurable functions $f: X \rightarrow Y$ and $g: Z \rightarrow W$, we define partial measurable functions $f + g: X + Z \rightarrow Y + W$ and $f \times g: X \times Z \rightarrow Y \times W$ by

$$\begin{aligned} (f + g)(\bullet, x) &= \begin{cases} (\bullet, y), & \text{if } f(x) \text{ is defined and is equal to } y, \\ \text{undefined}, & \text{otherwise,} \end{cases} \\ (f + g)(\circ, z) &= \begin{cases} (\circ, w), & \text{if } g(z) \text{ is defined and is equal to } w, \\ \text{undefined}, & \text{otherwise,} \end{cases} \\ (f \times g)(x, z) &= \begin{cases} (y, w), & \text{if } f(x) \text{ is defined and is equal to } y \\ & \text{and } g(w) \text{ is defined and is equal to } z, \\ \text{undefined}, & \text{otherwise.} \end{cases} \end{aligned}$$

We note that $(\mathbf{pMeas}, 1, \times)$ and $(\mathbf{pMeas}, \emptyset, +)$ are symmetric monoidal categories. We also note that $X \times (-)$ distributes over the coproducts, i.e., the canonical arrow

$$\text{dst}_{X,Y,Z} = [X \times \text{inl}_{Y,Z}, X \times \text{inr}_{Y,Z}]: X \times Y + X \times Z \rightarrow X \times (Y + Z)$$

is an isomorphism.

The notion of trace introduced by Joyal, Street and Verity [26] plays important role in this section.

Definition 5.3. Let $(\mathcal{C}, I, \otimes)$ be a symmetric monoidal category. A trace operator on $(\mathcal{C}, I, \otimes)$ is a family $\{\mathbf{tr}_{X,Y}^Z\}_{X,Y,Z \in \mathcal{C}}$ satisfying the following axioms:

- (Dinaturality) For all \mathcal{C} -arrows $f: X \otimes Z \rightarrow Y \otimes Z$, $g: X' \rightarrow X$ and $h: Y \rightarrow Y'$,

$$h \circ \mathbf{tr}_{X,Y}^Z(f) \circ g = \mathbf{tr}_{X',Y'}^Z((h \otimes Z) \circ f \circ (g \otimes Z)).$$

- (Sliding) For all \mathcal{C} -arrows $f: X \otimes Z \rightarrow Y \otimes W$, $g: W \rightarrow Z$,

$$\mathbf{tr}_{X,Y}^Z((Y \otimes g) \circ f) = \mathbf{tr}_{X,Y}^Z(f \circ (X \otimes g)).$$

- (Vanishing I) For all \mathcal{C} -arrows $f: X \otimes I \rightarrow Y \otimes I$,

$$\mathbf{tr}_{X,Y}^I(f) = f.$$

- (Vanishing II) For all \mathcal{C} -arrows $f: X \otimes Z \otimes W \rightarrow Y \otimes Z \otimes W$,

$$\mathbf{tr}_{X,Y}^{Z \otimes W}(f) = \mathbf{tr}_{X,Y}^Z(\mathbf{tr}_{X \otimes Z, Y \otimes Z}^W(f)).$$

- (Superposing) For all \mathcal{C} -arrows $f: X \otimes Z \rightarrow Y \otimes Z$,

$$W \otimes \mathbf{tr}_{X,Y}^Z(f) = \mathbf{tr}_{W \otimes X, W \otimes Y}^{W \otimes Z}(W \otimes f).$$

- (Yanking) For all $X \in \mathcal{C}$,

$$\mathbf{tr}_{X,X}^X(\sigma_{X,X}) = \text{id}_X$$

where $\sigma_{X,Y}: X \otimes Y \rightarrow Y \otimes X$ is the braiding.

A symmetric monoidal category $(\mathcal{C}, I, \otimes)$ endowed with a trace operator \mathbf{tr} is called a traced symmetric monoidal category.

We give a trace operator on $(\mathbf{pMeas}, \emptyset, +)$. The symmetric monoidal category $(\mathbf{pMeas}, \emptyset, +)$ is enriched over $\omega\mathbf{Cppo}$, which is the cartesian category of pointed ω -cpo's and continuous functions. The partial order on a hom-set $\mathbf{pMeas}(X, Y)$ is given by

$$f \leq g \iff \text{for all } x \in X, \text{ if } f(x) \text{ is defined, then } g(x) \text{ is defined, and } f(x) = g(x).$$

The least arrow $\perp_{X,Y}: X \rightarrow Y$ is the empty partial measurable function. The $\omega\mathbf{Cppo}$ -enrichment induces an iterator

$$\mathbf{iter}_{X,Y}: \mathbf{pMeas}(X, Y + X) \rightarrow \mathbf{pMeas}(X, Y)$$

given by

$$\mathbf{iter}_{X,Y}(f) = \text{the least fixed point of } \left(g: X \rightarrow Y \mapsto [\text{id}_Y, g] \circ f: X \rightarrow Y \right).$$

The operator \mathbf{iter} induces another operator

$$\mathbf{tr}_{X,Y}^Z: \mathbf{pMeas}(X + Z, Y + Z) \rightarrow \mathbf{pMeas}(X, Y)$$

given by

$$\mathbf{tr}_{X,Y}^Z(f) = [\text{id}_Y, \mathbf{iter}_{Z,Y}(f \circ \text{inr}_{X,Z})] \circ f \circ \text{inl}_{X,Z}.$$

Concretely, for a partial measurable function $f: X + Z \rightarrow Y + Z$,

$$\mathbf{tr}_{X,Y}^Z(f)(x) \text{ is defined and is equal to } y$$

if and only if either $f(\bullet, x) = (\bullet, y)$ or there is a finite sequence $z_1, \dots, z_n \in Z$ such that

$$f(\bullet, x) = (\circ, z_1), \quad f(\circ, z_1) = (\circ, z_2), \quad \dots \quad f(\circ, z_{n-1}) = (\circ, z_n), \quad f(\circ, z_n) = (\bullet, y).$$

Proposition 5.9. *The family of operators $\{\mathbf{tr}_{X,Y}^Z\}_{X,Y,Z \in \mathbf{pMeas}}$ is a trace operator of the symmetric monoidal category $(\mathbf{pMeas}, \emptyset, +)$. Furthermore, the trace operator is uniform [46] with respect to partial measurable functions : for all partial measurable functions $f: X + Z \rightarrow Y + Z$, $f: X + W \rightarrow Y + W$ and $h: Z \rightarrow W$, if*

$$\begin{array}{ccc} X + Z & \xrightarrow{f} & Y + Z \\ \downarrow X+h & & \downarrow Y+h \\ X + W & \xrightarrow{g} & Y + W \end{array}$$

commutes, then

$$\mathbf{tr}_{X,Y}^Z(f) = \mathbf{tr}_{X,Y}^W(g).$$

Proof. It is straightforward to adapt the argument in [47, Section A]. \square

We will use the next proposition to construct a trace operator for Mealy machines.

Proposition 5.10. *For any partial measurable function $f: X + Z \rightarrow Y + Z$ and a measurable space W ,*

$$W \times \mathbf{tr}_{X,Y}^Z(f) = \mathbf{tr}_{W \times X, W \times Y}^{W \times Z} (\text{dst}_{W,Y,Z}^{-1} \circ (W \times f) \circ \text{dst}_{W,X,Z}).$$

Proof. For any $w \in W$, we show that

$$\mathbf{tr}_{W \times X, W \times Y}^{W \times Z} (\text{dst}_{W,Y,Z}^{-1} \circ (W \times f) \circ \text{dst}_{W,X,Z}) \circ (w \times \text{id}_X) = w \times \mathbf{tr}_{X,Y}^Z(f)$$

where we identify w with the arrow from $1 = \{*\}$ to W that sends $*$ to w . Because

$$\text{dst}_{W,Y,Z}^{-1} \circ (W \times f) \circ \text{dst}_{W,X,Z} \circ (w \times X + w \times Z) = (w \times Y + w \times Z) \circ f,$$

it follows from uniformity that

$$\mathbf{tr}_{X, W \times Y}^{W \times Z} (\text{dst}_{W,Y,Z}^{-1} \circ (W \times f) \circ \text{dst}_{W,X,Z} \circ (w \times X + W \times Z)) = \mathbf{tr}_{W \times X, Y}^Z((w \times Y + Z) \circ f).$$

By dinaturality, we obtain

$$\mathbf{tr}_{W \times X, W \times Y}^{W \times Z} (\text{dst}_{W,Y,Z}^{-1} \circ (W \times f) \circ \text{dst}_{W,X,Z}) \circ (w \times \text{id}_X) = w \times \mathbf{tr}_{X,Y}^Z(f).$$

Since this is true for any $w \in W$, we see that $W \times \mathbf{tr}_{X,Y}^Z(f)$ is equal to

$$\mathbf{tr}_{W \times X, W \times Y}^{W \times Z} (\text{dst}_{W,Y,Z}^{-1} \circ (W \times f) \circ \text{dst}_{W,X,Z}).$$

\square

5.6.2 The Category of Mealy Machines

Definition 5.4. *We define a category **Mealy** by:*

- *objects are **Int**-objects X ; and*
- *arrows $f: X \multimap Y$ are behavioural equivalence classes of Mealy machines from X to Y .*

*We denote a wide subcategory of **Mealy** consisting of **Int**-object X such $X^- = \emptyset$ by **Mealy** $_+$.*

Intuitively, while arrows in **Mealy** are bidirectional Mealy machines, arrows in **Mealy** $_+$ are “one-way” Mealy machines. We consider the wide subcategory **Mealy** $_+$ because categorical structure of **Mealy** $_+$ is easier to describe than that of **Mealy**, and categorical structure of **Mealy** is induced by that of **Mealy** $_+$.

The identity arrow and the composition of **Mealy** $_+$ is given by the identity Mealy machine $[\text{id}_X]$ and the composition of Mealy machine:

$$[M] \circ [N] = [M \circ N].$$

Concrete description of the composition of Mealy machines between **Mealy** $_+$ -objects is easy: for **Mealy** $_+$ -objects X and Y , and for Mealy machines $M: X \multimap Y$ and $N: Y \multimap Z$, the composition $N \circ M: X \multimap Z$ consists of:

- $S_{N \circ M} = S_N \times S_M$;
- $s_{N \circ M} = (s_N, s_M)$;
- $\tau_{N \circ M}$ given by

$$\begin{aligned} (X^+ + \emptyset) \times S_g \times S_f &\xrightarrow{\cong} (X^+ + \emptyset) \times S_f \times S_g \xrightarrow{\tau_f \times S_g} \\ &(Y^+ + \emptyset) \times S_f \times S_g \xrightarrow{\cong} (Y^+ + \emptyset) \times S_g \times S_f \xrightarrow{\tau_g \times S_f} (Z^+ + \emptyset) \times S_g \times S_f. \end{aligned}$$

The composition of transition functions makes sense because $X^- = Z^- = \emptyset$. From this concrete description, it is easy to check that the composition of **Mealy**₊-arrows is well-defined. In fact, if $h: S_M \rightarrow S_{M'}$ realizes $M \preceq M'$ and $h': S_N \rightarrow S_{N'}$ realizes $N \preceq N'$, then $h' \times h$ realizes $N \circ M \preceq N' \circ M'$. Therefore, the symmetric transitive closure \simeq is compatible with the composition.

Proposition 5.11. *The category **Mealy**₊ with (\mathbf{l}, \otimes) is a symmetric monoidal category where the monoidal product of **Mealy**₊-arrows $[M]: X \multimap Y$ and $[N]: Z \multimap W$ is given by*

$$[M] \otimes [N] = [M \otimes N].$$

Proof. It is easy to see that objects in **Mealy**₊ are closed under the monoidal product of **Int**-objects. Thanks to simplicity of the composition of **Mealy**₊-arrows, we can easily check that the monoidal product of Mealy machines between **Mealy**₊-objects is compatible with behavioural equivalence and that $(\mathbf{Mealy}_+, I, \otimes)$ is a symmetric monoidal category. \square

Furthermore, **Mealy**₊ inherits the trace operator of **pMeas**. For a **Mealy**₊-arrow $[M]: X \otimes Z \multimap Y \otimes Z$, we define a **Mealy**₊-arrow $\text{Tr}_{X,Y}^Z[M]: X \multimap Y$ to be the equivalence class of a Mealy machine $N: X \multimap Y$ given by

$$S_N = S_M, \quad s_N = s_M$$

and

$$\tau_N = \mathbf{tr}_{(X^+ + \emptyset) \times S_f, (Y^+ + \emptyset) \times S_f}^{(Z^+ + \emptyset) \times S_f} \left(\begin{array}{c} (X^+ + \emptyset) \times S_f + (Z^+ + \emptyset) \times S_f \\ \downarrow \cong \\ ((X^+ + Z^+) + \emptyset) \times S_f \\ \downarrow \tau_M \\ ((Y^+ + Z^+) + \emptyset) \times S_f \\ \downarrow \cong \\ (Y^+ + \emptyset) \times S_f + (Z^+ + \emptyset) \times S_f \end{array} \right).$$

Proposition 5.12. *The family of operators $\{\text{Tr}_{X,Y}^Z\}_{X,Y,Z \in \mathbf{Mealy}_+}$ is a trace operator on the symmetric monoidal category $(\mathbf{Mealy}_+, \mathbf{l}, \otimes)$.*

Proof. Well-definedness of $\text{Tr}_{X,Y}^Z(-)$ follows from uniformity of the trace operator on **pMeas**. Sliding, vanishing I, vanishing II, superposing and yanking for Tr follow from that of \mathbf{tr} . Dinaturality for Tr follows from dinaturality of \mathbf{tr} and Proposition 5.10. \square

We recall the notions of **Int-construction** [26] and *compact closed category*.

Definition 5.5 (Int-construction). *Let $(\mathcal{C}, I, \otimes, \mathbf{tr})$ be a traced symmetric monoidal category. We define a category **Int**(\mathcal{C}) by:*

- *objects are pairs (X^+, X^-) of \mathcal{C} -objects;*
- *arrows from (X^+, X^-) to (Y^+, Y^-) are \mathcal{C} -arrows from $X^+ \otimes Y^-$ to $Y^+ \otimes X^-$.*

The identity on (X^+, X^-) is given by the identity on $X^+ \otimes X^-$, and the composition of $\mathbf{Int}(\mathcal{C})$ -arrows $f: (X^+, X^-) \rightarrow (Y^+, Y^-)$ and $g: (Y^+, Y^-) \rightarrow (Z^+, Z^-)$ is given by

$$\mathrm{tr}_{X^+ \otimes Z^-, Z^+ \otimes X^-}^{Y^-} \left((X^+ \otimes \sigma_{Z^-, Y^-}) \circ (f \otimes Z^-) \circ (Y^+ \otimes \sigma_{X^-, Z^-}) \circ (g \otimes X^-) \circ (Z^+ \otimes \sigma_{Y^-, X^-}) \right).$$

Here, we omit some coherence isomorphisms.

Definition 5.6. A compact closed category is a symmetric monoidal category $(\mathcal{C}, I, \otimes)$ with a function $(-)^{\perp}: \mathrm{obj}(\mathcal{C}) \rightarrow \mathrm{obj}(\mathcal{C})$ and families of \mathcal{C} -arrows

$$\{\eta_X: I \rightarrow X \otimes X^{\perp}\}_{X \in \mathcal{C}}, \quad \{\epsilon_X: X^{\perp} \otimes X \rightarrow I\}_{X \in \mathcal{C}}$$

such that

$$(\epsilon_X \otimes X^{\perp}) \circ (X^{\perp} \otimes \eta_X) = \mathrm{id}_{X^{\perp}}, \quad (X \otimes \epsilon_X) \circ (\eta_X \otimes X) = \mathrm{id}_X.$$

For $X \in \mathcal{C}$, the object X^{\perp} is called the dual object of X .

Theorem 5.3 ([26]). The category $\mathbf{Int}(\mathcal{C})$ is a compact closed category. The unit and the monoidal product are given by

$$(I, I), \quad (X^+, X^-) \otimes (Y^+, Y^-) = (X^+ \otimes Y^+, Y^- \otimes X^-).$$

The dual object of (X^+, X^-) is (X^-, X^+) . The unit arrow $\eta_{(X^+, X^-)}$ and the counit arrow $\epsilon_{(X^+, X^-)}$ are given by

$$\eta_{(X^+, X^-)} = \mathrm{id}_{X^+ \otimes X^-}, \quad \epsilon_{(X^+, X^-)} = \mathrm{id}_{X^- \otimes X^+}.$$

Corollary 5.1. The category \mathbf{Mealy} is a compact closed category. The monoidal structure is given by (I, \otimes) , and the unit and the counit are given by unit_X and counit_X respectively.

Proof. It is straightforward to check that \mathbf{Mealy} is isomorphic to $\mathbf{Int}(\mathbf{Mealy}_+)$, and the compact closed structure is given by data provided in Section 5. \square

6 Mealy Machine Semantics for PCFSS

We interpret a type A as the \mathbf{Int} -object $\llbracket A \rrbracket$ given by

$$\llbracket \mathbf{Unit} \rrbracket = I, \quad \llbracket \mathbf{Real} \rrbracket = R, \quad \llbracket A \rightarrow B \rrbracket = S \otimes !\llbracket B \rrbracket \otimes !\llbracket A \rrbracket^{\perp}.$$

We define interpretation of contexts by

$$\llbracket x : A, \dots, y : B \rrbracket = \llbracket A \rrbracket \otimes \dots \otimes \llbracket B \rrbracket.$$

When Δ is the empty sequence, we define $\llbracket \Delta \rrbracket$ to be I .

For interpreting conditional branching, we use the following proposition.

Proposition 6.1. For any type A , there is a partial measurable embedding $e: S + N \times \llbracket A \rrbracket^{-} \rightarrow S$.

Proof. We first define an embedding from $\llbracket A \rrbracket^{-}$ to S by induction on A . We note that for any type A , we have $\llbracket A \rrbracket^{+} = \llbracket A \rrbracket^{-}$. Base cases are easy. For induction step,

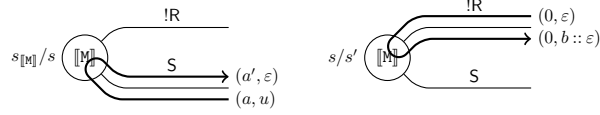
$$\begin{aligned} \llbracket A \rightarrow B \rrbracket^{-} &= S + N \times \llbracket B \rrbracket^{-} + N \times \llbracket A \rrbracket^{-} \subseteq S + N \times S + N \times S && (\text{induction hypothesis}) \\ &\subseteq S + R \times S + R \times S && (N \subseteq R) \\ &\subseteq S + S + S && (R \times S \subseteq S) \\ &\cong \{0, 1, 2\} \times S \\ &\subseteq R \times S \\ &\subseteq S. \end{aligned}$$

The statement follows from $S + N \times \llbracket A \rrbracket^{-} \subseteq \llbracket \mathbf{Unit} \rightarrow A \rrbracket^{-}$. \square

- for $(a, u) \in \mathbb{R}_{\geq 0} \times \mathbb{T}$, if there are $s, s' \in S_{\llbracket \mathbf{M} \rrbracket}$ such that

$$\begin{aligned}\tau_{\llbracket \mathbf{M} \rrbracket}((\circ, (\circ, (a, u))), s_{\llbracket \mathbf{M} \rrbracket}) &= ((\bullet, (\bullet, (a', \varepsilon))), s), \\ \tau_{\llbracket \mathbf{M} \rrbracket}((\circ, (\bullet, (0, \varepsilon))), s) &= (\bullet, (\circ, (0, b :: \varepsilon)), s'),\end{aligned}$$

i.e., if we have the following transitions:



then we define $\mathfrak{o}(\mathbf{M})(a, u)$ to be (a', b) ;

- otherwise, $\mathfrak{o}(\mathbf{M})(a, u)$ is undefined.

Theorem 7.1 (Adequacy). *For any closed term $\vdash \mathbf{M} : \mathbf{Real}$ and for any $(a, u) \in \mathbb{R}_{\geq 0} \times \mathbb{T}$, we have*

$$\mathfrak{o}(\mathbf{M})(a, u) = (a', b) \iff (\mathbf{M}, a, u) \rightarrow^* (b, a', \varepsilon).$$

Corollary 7.1. *For any closed term $\vdash \mathbf{M} : \mathbf{Real}$, partial functions $\mathbf{weight}(\mathbf{M}) : \mathbb{R}_{\geq 0} \times \mathbb{T} \rightarrow \mathbb{R}_{\geq 0}$ and $\mathbf{val}(\mathbf{M}) : \mathbb{R}_{\geq 0} \times \mathbb{T} \rightarrow \mathbb{R}$ given as follows*

$$(\mathbf{weight}(\mathbf{M})(a, u), \mathbf{val}(\mathbf{M})(a, u)) = \begin{cases} (a', b), & \text{if } (\mathbf{M}, a, u) \rightarrow^* (b, a', \varepsilon), \\ \text{undefined}, & \text{otherwise} \end{cases}$$

are partial measurable functions.

7.2 Distribution-Based Operational Semantics

For a closed term $\mathbf{M} : \mathbf{Real}$, we define measurable functions $\mathfrak{o}_0(\mathbf{M}) : \mathbb{T} \rightarrow \mathbb{R}_{\geq 0}$ and $\mathfrak{o}_1(\mathbf{M}) : \mathbb{T} \rightarrow \mathbb{R}$ by

$$\begin{aligned}\mathfrak{o}_0(\mathbf{M})(u) &= \begin{cases} a, & \text{if } \exists b \in \mathbb{R}, \mathfrak{o}(\mathbf{M})(1, u) = (a, b), \\ 0, & \text{otherwise,} \end{cases} \\ \mathfrak{o}_1(\mathbf{M})(u) &= \begin{cases} b, & \text{if } \exists a \in \mathbb{R}_{\geq 0}, \mathfrak{o}(\mathbf{M})(1, u) = (a, b), \\ 0, & \text{otherwise.} \end{cases}\end{aligned}$$

Then we define a measure $\mathfrak{D}(\mathbf{M})$ on \mathbb{R} by:

$$\mathfrak{D}(\mathbf{M})(A) = \sum_{n \in \mathbb{N}} \int_{\mathbb{R}_{[0,1]}^n} \mathfrak{o}_0(\mathbf{M})(u) [\mathfrak{o}_1(\mathbf{M})(u) \in A] \, du.$$

Theorem 7.2 (Adequacy). *For any closed term $\vdash \mathbf{M} : \mathbf{Real}$, we have $\mathbf{M} \Rightarrow_{\infty} \mathfrak{D}(\mathbf{M})$.*

It follows from our adequacy theorems that sampling-based operational semantics induces distribution-based operational semantics.

Corollary 7.2. *For any closed term $\vdash \mathbf{M} : \mathbf{Real}$,*

$$\mathbf{M} \Rightarrow_{\infty} \sum_{n \in \mathbb{N}} \int_{\mathbb{R}_{[0,1]}^n} \mathbf{weight}(\mathbf{M})(u) [\mathbf{val}(\mathbf{M})(u) \in A] \, du.$$

A result analogous to Corollary 7.2 has already been proved by way of a purely operational (and quite laborious) argument in an untyped setting in which, however, **score** is not available in its full generality [28]. Here, it is just an easy corollary of our adequacy theorems.

8 Proof of Adequacy Theorems

Lemma 8.1. *For any term $\Delta, \mathbf{x} : \mathbf{A} \vdash \mathbf{M} : \mathbf{B}$ and for any closed value $\vdash \mathbf{V} : \mathbf{A}$,*

$$\llbracket \mathbf{M} \rrbracket \circ (\llbracket \Delta \rrbracket \otimes !(\mathbf{V})) \simeq \llbracket \mathbf{M}\{\mathbf{V}/\mathbf{x}\} \rrbracket.$$

Proof. By induction on \mathbf{M} . □

Lemma 8.2. *For all closed terms $\mathbf{M}, \mathbf{N} : \mathbf{A}$, if $\mathbf{M} \xrightarrow{\text{red}} \mathbf{N}$, then $\llbracket \mathbf{M} \rrbracket = \llbracket \mathbf{N} \rrbracket$.*

Proof. By case analysis. For the case of recursion, see Proposition 9.2 in Section 9. □

We first prove soundness.

Proposition 8.1. *For any closed term $\mathbf{M} : \mathbf{Real}$ and for any $(a, u) \in \mathbb{R}_{\geq 0} \times \mathbb{T}$, if $(\mathbf{M}, a, u) \rightarrow^* (b, a', \varepsilon)$, then $\mathfrak{o}(\mathbf{M})(a, u) = (a', b)$.*

Proof. By induction on the length of \rightarrow^* . (Base case) Easy. (Induction step) By case analysis on the first evaluation step of $(\mathbf{M}, a, u) \rightarrow^* (b, a', \varepsilon)$.

- If the first evaluation step is of the form $(\mathbf{E}[\mathbf{N}], a, u) \rightarrow (\mathbf{E}[\mathbf{L}], a', u')$ for some $\mathbf{N} \xrightarrow{\text{red}} \mathbf{L}$, then by Lemma 8.2, we have $\mathbf{E}[\mathbf{N}] = \mathbf{E}[\mathbf{L}]$. Because $(\mathbf{E}[\mathbf{L}], a', u') \rightarrow^* (b, a', \varepsilon)$, by induction hypothesis, we obtain $\mathfrak{o}(\mathbf{E}[\mathbf{L}])(a', u') = (a', b)$. Hence, $\mathfrak{o}(\mathbf{E}[\mathbf{N}])(a, u) = \mathfrak{o}(\mathbf{E}[\mathbf{L}], a', u') = (a', b)$.
- If the first evaluation step is of the form $(\mathbf{E}[\mathbf{score}(\mathbf{r}_c)], a, u) \rightarrow (\mathbf{E}[\mathbf{skip}], |c| a, u)$, then by induction hypothesis, we have $\mathfrak{o}(\mathbf{E}[\mathbf{skip}])(|c| a, u) = (a', b)$. Therefore, by the definition of the Mealy machine \mathbf{sc} , we see that $\mathfrak{o}(\mathbf{E}[\mathbf{score}(\mathbf{r}_c)])(a, u)$ is (a', b) .
- If the first evaluation step is of the form $(\mathbf{E}[\mathbf{sample}], a, c :: u) \rightarrow (\mathbf{E}[\mathbf{r}_c], a, u)$, then by induction hypothesis, $\mathfrak{o}(\mathbf{E}[\mathbf{r}_c])(a, u)$ is (a', b) . Therefore, by the definition of the Mealy machine \mathbf{sample} , we see that $\mathfrak{o}(\mathbf{E}[\mathbf{sample}])(a, c :: u)$ is (a', b) .

□

It remains to prove that $\mathfrak{o}(\mathbf{M})(a, u) = (a', b)$ implies that $(\mathbf{M}, a, u) \rightarrow^* (b, a', \varepsilon)$. We use logical relations. We define a binary relation O between closed terms of type \mathbf{Real} and Mealy machines from \mathbf{I} to $\mathbf{S} \otimes \mathbf{!R}$ by

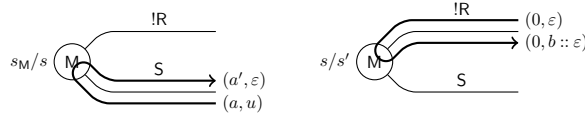
$$(\mathbf{M}, \mathbf{M}) \in O \iff \text{if } \mathfrak{o}(\mathbf{M})(a, u) = (a', b), \text{ then } (\mathbf{M}, a, u) \rightarrow^* (b, a', \varepsilon)$$

where $\mathfrak{o}(\mathbf{M}) : \mathbb{R}_{\geq 0} \times \mathbb{T} \rightarrow \mathbb{R}_{\geq 0} \times \mathbb{R}$ is a partial measurable function given by: for each $(a, u) \in \mathbb{R}_{\geq 0} \times \mathbb{T}$,

- if there are states $s, s' \in S_{\mathbf{M}}$ such that

$$\begin{aligned} \tau_{\mathbf{M}}((\circ, (\circ, (a, u))), s_{\mathbf{M}}) &= ((\bullet, (\bullet, (a', \varepsilon))), s), \\ \tau_{\mathbf{M}}((\circ, (\bullet, (0, \varepsilon))), s) &= (\bullet, (\circ, (0, b :: \varepsilon)), s'), \end{aligned}$$

i.e., if we have the following transitions:



then we define $\mathfrak{o}(\mathbf{M})(a, u)$ to be (a', b) ;

- otherwise, $\mathfrak{o}(\mathbf{M})(a, u)$ is undefined.

We then inductively define binary relations

$$\begin{aligned} R_A &\subseteq \{\text{closed values of type } A\} \times \{\text{Mealy machines from } I \text{ to } \llbracket A \rrbracket\} \\ R_A^\top &\subseteq \{\text{evaluation contexts } x : A \vdash E[x] : \text{Real}\} \times \{\text{Mealy machines from } \llbracket A \rrbracket \text{ to } S \otimes !R\} \\ \bar{R}_A &\subseteq \{\text{closed terms of type } A\} \times \{\text{Mealy machines from } I \text{ to } S \otimes \llbracket A \rrbracket\} \end{aligned}$$

by

$$\begin{aligned} R_{\text{Real}} &= \{(r_a, r_a) : a \in \mathbb{R}\}, \\ R_{\text{Unit}} &= \{(\text{skip}, \text{id}_I)\}, \\ R_{A \rightarrow B} &= \{(V, M) : \forall (W, N) \in R_A, (VW, (S \otimes \llbracket B \rrbracket \otimes \text{counit}_{\llbracket A \rrbracket}) \circ (M \otimes !N)) \in \bar{R}_B\}, \\ R_A^\top &= \{(E[-], E) : \forall (V, M) \in R_A, (E[V], E \circ !M) \in O\}, \\ \bar{R}_A &= \{(M, M) : \forall (E[-], E) \in R_A^\top, (E[M], (m \otimes !R) \circ (S \otimes E) \circ M) \in O\} \end{aligned}$$

We list some properties of the logical relations.

Lemma 8.3. *Let A be a type.*

1. If $(V, M) \in R_A$, then $(V, e \otimes !M) \in \bar{R}_A$.
2. If $(M, M) \in \bar{R}_A$ and $N \xrightarrow{\text{red}} M$, then $(N, M) \in \bar{R}_A$.
3. If $(M, M) \in \bar{R}_A$ and $M \xrightarrow{\text{red}} N$, then $(N, M) \in \bar{R}_A$.
4. If $(M, M) \in \bar{R}_A$ and $M \simeq N$, then $(M, N) \in \bar{R}_A$.
5. For any closed term $M : A$, $(M, \text{bot}_{S \otimes \llbracket A \rrbracket}) \in \bar{R}_A$ where $\text{bot}_X : I \multimap X$ is a token machine whose transition function is the empty partial measurable function.
6. For any closed value $V : A \rightarrow B$, $(V, \text{bot}_{\llbracket A \rightarrow B \rrbracket}) \in R_{A \rightarrow B}$.
7. If $(M, M_i) \in \bar{R}_A$ and $[M_1] \leq [M_2] \leq \dots$, then $(M, N) \in \bar{R}_A$ where $[N]$ is the least upper bound of the ω -chain $[M_1] \leq [M_2] \leq \dots$.

Proof. We can check these items by unfolding the definition of O and the logical relations. \square

Lemma 8.4 (Basic Lemma). *Let $\Delta = (x : A_1, \dots, x_n : A_n)$ be a context.*

- For any term $\Delta \vdash M : A$ and for any $(V_i, N_i) \in R_{A_i}$ for $i = 1, 2, \dots, n$, we have

$$(M\{V_1/x_1, \dots, V_n/x_n\}, \llbracket M \rrbracket \circ (!N_1 \otimes \dots \otimes !N_n)) \in \bar{R}_A.$$

- For any value $\Delta \vdash V : A$ and for any $(V_i, N_i) \in R_{A_i}$ for $i = 1, 2, \dots, n$, we have

$$(V\{V_1/x_1, \dots, V_n/x_n\}, \llbracket V \rrbracket \circ (!N_1 \otimes \dots \otimes !N_n)) \in R_A.$$

Proof. By induction on M and V . Most cases follow from Lemma 8.3. For $M = \text{sample}$ and $M = \text{score}(V)$, we check the statement by unfolding the definition of sa and sc . Here, we only check for $M = \text{sample}$ and $M = \text{fix}_{A,B}(f, x, N)$.

- When $M = \text{sample}$, for any (E, E) in R_{Real}^\top , if

$$o((m \otimes !R) \circ (S \otimes E) \circ \text{sa})(a, u) = (a', b),$$

then by the definition of sa , we see that $u = c :: u'$ for some $c \in \mathbb{R}_{[0,1]}$ and $u' \in \mathbb{T}$ such that

$$o(E \circ !r_c)(a, u') = (a', b).$$

Because $(E, E) \in R_{\text{Real}}^\top$, we obtain $(E[r_c], a, u') \rightarrow^* (b, a', \varepsilon)$. Hence,

$$(E[\text{sample}], a, u) \rightarrow^* (b, a', \varepsilon).$$

- When $M = \text{fix}_{A,B}(\mathbf{f}, \mathbf{x}, N)$, for simplicity, we suppose that M is a closed term. By induction hypothesis, we can check that

$$(M, \llbracket \lambda \mathbf{x}^A. N \rrbracket \circ !\llbracket \lambda \mathbf{x}^A. N \rrbracket \circ \dots \circ !^k \llbracket \lambda \mathbf{x}^A. N \rrbracket \circ \text{bot}_{I, !^k \llbracket A \rightarrow B \rrbracket}) \in R_{A \rightarrow B}$$

by induction on n . Because $\llbracket M \rrbracket$ is the least upper bound of $\llbracket \lambda \mathbf{x}^A. M \rrbracket \circ !\llbracket \lambda \mathbf{x}^A. N \rrbracket \circ \dots \circ !^k \llbracket \lambda \mathbf{x}^A. N \rrbracket \circ \text{bot}_{I, !^k \llbracket A \rightarrow B \rrbracket}$ (Proposition 9.1), we obtain $(M, \llbracket M \rrbracket) \in R_{A \rightarrow B}$ by Lemma 8.3.

□

Theorem 8.1. *For any closed term $\vdash M : \text{Real}$ and for any $(a, u) \in \mathbb{R}_{\geq 0} \times \mathbb{T}$, we have*

$$\mathfrak{o}(M)(a, u) = (a', b) \iff (M, a, u) \rightarrow^* (b, a', \varepsilon).$$

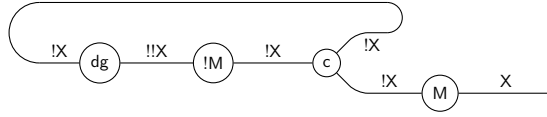
Proof. If $(M, a, u) \rightarrow^* (b, a', \varepsilon)$, then we have $\mathfrak{o}(M)(a, u) = (a', b)$ by Proposition 8.1. If $\mathfrak{o}(M)(a, u) = (a', b)$, then because $([-], \mathbf{e} \otimes \text{id}_{!R})$ is an element of R_{Real}^\top , we obtain $(M, a, u) \rightarrow^* (b, a', \varepsilon)$ by Lemma 8.4. □

9 Approximation Lemma

Let $M : !X \multimap X$ be a Mealy machine. In this section, we give a way to calculate a Mealy machine $M^\dagger : I \rightarrow !X$ given by

$$M^\dagger = (M \otimes \text{cunit}_{!X}) \circ ((c_X \circ !M \circ \text{dg}_X) \otimes \text{id}_{!X}) \circ \text{unit}_{!X}$$

Diagrammatically, M^\dagger consists of digging, contraction and a feed back loop:



This construction already appeared in the interpretation of the fixed point operator. In fact, for a term $\mathbf{f} : A \rightarrow B, \mathbf{x} : A \vdash M : B$, we have $\llbracket \text{fix}_{A,B}(\mathbf{f}, \mathbf{x}, M) \rrbracket = \llbracket \lambda \mathbf{x}^A. M \rrbracket^\dagger$.

The goal of this section is to show that M^\dagger is a fixed point of M and can be approximated by a family of Mealy machines

$$M \circ \text{bot}_{I, !X}, M \circ !(M \circ \text{bot}_{I, !X}), M \circ !(M \circ !(M \circ \text{bot}_{I, !X})), \dots : I \multimap X.$$

9.0.1 Parametrized Modal Operator and Parametrized Loop Operator

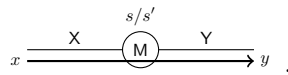
We introduce parametrization of the modal operator $!$. For a subset $\alpha \subseteq \mathbb{N}$ and for a Mealy machine $M : X \multimap Y$, we define a Mealy machine $!_\alpha M : !X \multimap !Y$ by: the state space and the initial state of $!_\alpha M$ are given by

$$S_{!_\alpha M} = S_{!M} = S_M^\mathbb{N}, \quad s_{!_\alpha M} = s_{!M}$$

and $\tau_{!_\alpha M}$ is given by

$$\tau_{!M}((i, (n, z)), (s_n)_{n \in \mathbb{N}}) = \begin{cases} ((j, (n, w)), (s_1, \dots, s_{n-1}, t, s_{n+1}, \dots)), & \text{if } n \in \alpha \text{ and } \tau_M((i, z), s_n) = ((j, w), t), \\ \text{undefined}, & \text{otherwise} \end{cases}$$

where $i, j \in \{0, 1\}$ and z, w vary over the corresponding sets. For example, if we have



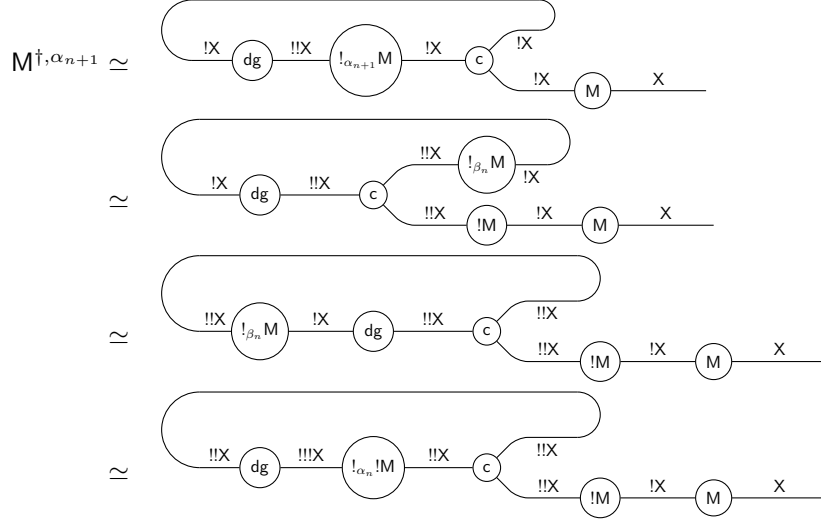


Figure 5: A Diagrammatic Proof of Lemma 9.2

then for any $t_1, t_2, \dots \in S_M$ and for any $n \in \mathbb{N}$, we have

$$\begin{array}{c} (t_1, \dots, t_{n-1}, s, t_{n+1}, \dots) / (t_1, \dots, t_{n-1}, s', t_{n+1}, \dots) \\ \xrightarrow{(n, x)} \text{(!M)} \xrightarrow{(n, y)} \end{array}$$

whenever $n \in \alpha$. When $n \notin \alpha$, there is no output from $!_\alpha M$. We can think $!_\alpha M$ as a “restriction” of $!M$ to α . In fact, $!M$ is equal to $!_{\mathbb{N}} M$.

We are interested in restrictions of $!$ to subsets $\alpha_n, \beta_n \subseteq \mathbb{N}$ inductively given by

$$\alpha_0 = \emptyset, \quad \beta_n = \{\langle i, j \rangle : i \in \alpha_n \text{ and } j \in \mathbb{N}\}, \quad \alpha_{n+1} = \{2i : i \in \mathbb{N}\} \cup \{2i + 1 : i \in \beta_n\}.$$

The definition of α_n and β_n are motivated by the following lemma.

Lemma 9.1. *For any $n \in \mathbb{N}$ and for any $M: X \multimap Y$, we have*

$$c_Y \circ !_{\alpha_{n+1}} M \simeq (!M \otimes !_{\beta_n} M) \circ c_X, \quad dg_Y \circ !_{\beta_{n+1}} M \simeq !_{\alpha_n} !M \circ dg_X.$$

By means of $!_\alpha$, we can also parametrize the operator $(-)^{\dagger}$. For $\alpha \subseteq \mathbb{N}$, and for $M: !X \multimap X$, we define $M^{\dagger, \alpha}: I \rightarrow !X$ by

$$M^{\dagger, \alpha} = (M \otimes \text{counit}_{!X}) \circ ((c_X \circ !_\alpha M) \circ dg_X) \otimes \text{id}_{!X} \circ \text{unit}_{!X}.$$

It is easy to see that M^{\dagger} is equal to $M^{\dagger, \mathbb{N}}$.

Lemma 9.2. *For any Mealy machine $M: !X \multimap X$, we have*

$$M^{\dagger, \alpha_{n+1}} \simeq M \circ (!M)^{\dagger, \alpha_n}.$$

Proof. See Figure 5. □

Lemma 9.3. *For any Mealy machine $M: !X \multimap X$ and for any $n \in \mathbb{N}$,*

$$(!M)^{\dagger, \alpha_n} \simeq !(M^{\dagger, \alpha_n}).$$

Proof. We prove the statement by induction on n . The base case follows from that the transition functions of $(!M)^{\dagger, \alpha_0}$ and $!(M^{\dagger, \alpha_0})$ are equal to the empty partial function. We next check the induction step. We have

$$\begin{aligned}
(!M)^{\dagger, \alpha_{n+1}} &\simeq !M \circ (!M)^{\dagger, \alpha_n} && \text{(Lemma 9.2)} \\
&\simeq !M \circ !(M)^{\dagger, \alpha_n} && \text{(Induction hypothesis)} \\
&\simeq !(M \circ (M)^{\dagger, \alpha_n}) && \text{(Functoriality)} \\
&\simeq !(M^{\dagger, \alpha_{n+1}}). && \text{(Lemma 9.2)}
\end{aligned}$$

□

Proposition 9.1. *For a Mealy machine $M: !X \multimap X$, we inductively define $\text{iter}_n(M): ! \multimap X$ by*

$$\text{iter}_0(M) = \text{bot}_{!X}, \quad \text{iter}_{n+1}(M) = M \circ !(\text{iter}_n(M)).$$

For all $n \in \mathbb{N}$, we have

$$S_{M^\dagger} = S_{M^\dagger, \alpha_n}, \quad s_{M^\dagger} = s_{M^\dagger, \alpha_n}, \quad M^{\dagger, \alpha_n} \simeq \text{iter}_n(M),$$

and

$$\tau_{M^\dagger, \alpha_0} \leq \tau_{M^\dagger, \alpha_1} \leq \tau_{M^\dagger, \alpha_2} \leq \dots, \quad \tau_{M^\dagger} = \bigvee_{n \geq 0} \tau_{M^\dagger, \alpha_n}.$$

Hence, we have an ascending chain

$$[M^{\dagger, \alpha_0}] \leq [M^{\dagger, \alpha_1}] \leq [M^{\dagger, \alpha_2}] \leq \dots,$$

and $[M^\dagger]$ is the least upper bound of the ascending chain $[\text{iter}_n(M)]$.

Proof. It follows from the definition of $!_{\alpha_n}$, we have

$$S_{!M} = S_{!_{\alpha_n} M}, \quad s_{!M} = s_{!_{\alpha_n} M}$$

for all $n \in \mathbb{N}$, and

$$\tau_{!_{\alpha_0} M} \leq \tau_{!_{\alpha_1} M} \leq \tau_{!_{\alpha_2} M} \leq \dots, \quad \tau_{!M} = \bigvee_{n \geq 0} \tau_{!_{\alpha_n} M}$$

Hence, by the definition of the composition and the monoidal product, we have

$$S_{M^\dagger} = S_{M^\dagger, \alpha_n}, \quad s_{M^\dagger} = s_{M^\dagger, \alpha_n}$$

for all $n \in \mathbb{N}$, and

$$\tau_{M^\dagger, \alpha_0} \leq \tau_{M^\dagger, \alpha_1} \leq \tau_{M^\dagger, \alpha_2} \leq \dots, \quad \tau_{M^\dagger} = \bigvee_{n \geq 0} \tau_{M^\dagger, \alpha_n}.$$

It remains to check $\text{iter}_n(M) \simeq M^{\dagger, \alpha_n}$. We show this by induction on n . For the base case, we have $M^{\dagger, \emptyset} \simeq \text{iter}_0(M)$ because these Mealy machines $M^{\dagger, \emptyset}$ and $\text{iter}_0(M)$ are behaviorally equivalent to $\text{bot}_{!X}$. For the induction step,

$$M \circ !(\text{iter}_n(M)) \stackrel{\text{induction hypothesis}}{\simeq} M \circ !(M^{\dagger, \alpha_n}) \stackrel{\text{Lemma 9.3}}{\simeq} M \circ (!M)^{\dagger, \alpha_n} \stackrel{\text{Lemma 9.2}}{\simeq} M^{\dagger, \alpha_{n+1}}.$$

Because $\text{iter}_{n+1}(M)$ is equal to $M \circ !(\text{iter}_n(M))$, we obtain $\text{iter}_{n+1}(M) \simeq M^{\dagger, \alpha_{n+1}}$. □

Proposition 9.2. *For any Mealy machine $M: !X \multimap X$,*

$$M \circ !(M^\dagger) \simeq M^\dagger.$$

Proof. Because $!(-)$ and the composition of Mealy machines are continuous, we have

$$M \circ !(M^\dagger) \simeq M \circ \bigvee_{n \in \mathbb{N}} !(M^{\dagger, \alpha_n}) \simeq \bigvee_{n \in \mathbb{N}} (M \circ !(M^{\dagger, \alpha_n})) \simeq \bigvee_{n \in \mathbb{N}} \text{iter}_{n+1}(M) \simeq \bigvee_{n \in \mathbb{N}} M^{\dagger, \alpha_{n+1}} \simeq M^\dagger.$$

□

10 How About S-Finite Kernels?

The reader experienced with the semantics of probabilistic programming languages have probably already wondered whether a GoI model for **PCFSS** could be given out of s-finite kernels instead of measurable functions, following Staton's work on the semantics of a first-order probabilistic programming language [32].

The answer is indeed positive: the kind of construction we have presented in Section 5 can in fact be adapted to the category of measurable spaces *and s-finite kernels*. The latter, being traced monoidal, has all the necessary structure one needs [27]. What one obtains proceeding this way is indeed a GoI model, but adequate only for the distribution-based operational semantics.

The interpretation of any program in this alternative GoI can be seen as structurally identical to the one from Section 5 once the sample and score operators are interpreted as usual, namely as those s-finite kernels which actually perform sampling and scoring *internally*. Below, we first recall the definition of s-finite kernel, and then we introduce Mealy machines whose transition is described in terms of an s-finite kernel, and we give some basic Mealy machines. Finally, we give an adequate GoI model for the distribution-based operational semantics.

Being adequate for the distribution-based semantics directly (and not by way of integration as in Theorem 7.2) has the pleasant consequence of validating a number of useful program transformations, and in particular commutation of sampling and scoring effects, see [28] for a thorough discussion about this topic, and about how s-finite kernels are a particularly nice way of achieving commutativity in presence of scoring.

10.1 S-finite Kernels

Let $k: X \rightsquigarrow Y$ be a kernel. We say that k is *finite* when there is a real number $c > 0$ such that for all $x \in X$ and $A \in \Sigma_Y$, we have $k(x, A) < c$. An *s-finite kernel* is a kernel $k: X \rightsquigarrow Y$ such that there is a countable family $\{k_n: X \rightsquigarrow Y\}$ of finite kernels such that $k(x, A) = \sum_{n \in \mathbb{N}} k_n(x, A)$ for all $x \in X$ and $A \in \Sigma_Y$. It is easy to see that s-finite kernels are closed under the pointwise addition. We write $\sum_{i \in I} k_i: X \rightsquigarrow Y$ for the pointwise addition of s-finite kernels $k_i: X \rightsquigarrow Y$. A *(sub)probability kernel* is a kernel $k: X \rightsquigarrow Y$ such that $k(x, -)$ is a (sub)probability measure on Y for all $x \in X$. Every (sub)probability kernel is a finite kernel.

Every measurable function $f: X \rightarrow Y$ gives rise to a probability kernel $\hat{f}: X \rightsquigarrow Y$ given by

$$\hat{f}(x, A) = [f(x) \in A].$$

We denote the probability kernel induced by the identity measurable function by $\text{id}_X: X \rightsquigarrow X$. Concretely, this is given by $\text{id}_X(x, A) = [x \in A]$.

We recall two constructions of s-finite kernels.

- (Composition) For s-finite kernels $k: X \rightsquigarrow Y$ and $h: Y \rightsquigarrow Z$, we define an s-finite kernel $h \circ k: X \rightsquigarrow Z$ by

$$(h \circ k)(x, C) = \int h(y, C) k(x, dy).$$

The composition of s-finite kernels is associative and satisfies the unit laws, namely, we have $k \circ \text{id}_X = k$ and $\text{id}_Y \circ k = k$.

- (Tensor product) For s-finite kernels $k: X \rightsquigarrow Y$ and $h: Z \rightsquigarrow W$, we define an s-finite kernel $k \otimes h: X \times Z \rightsquigarrow Y \times W$ to be the unique s-finite kernel such that for all $(x, z) \in X \times Z$ and for all $A \in \Sigma_Y$ and $B \in \Sigma_W$,

$$(k \otimes h)((x, z), A \times B) = k(x, A) h(z, B).$$

The tensor product and the coproduct of s-finite kernels is functorial. This means that these constructors are compatible with the composition and preserve identities. The following proposition summarizes categorical status of these structures.

Proposition 10.1. *The category of measurable spaces and s-finite kernels with \otimes forms a symmetric monoidal category where the unit object is 1. The object \emptyset is the zero object, and $X + Y$ with*

$$\widehat{\text{inl}}_{X,Y}: X \rightsquigarrow X + Y, \quad \widehat{\text{inr}}_{X,Y}: Y \rightsquigarrow X + Y,$$

forms the coproduct of X and Y where $\text{inl}_{X,Y}: X \rightarrow X + Y$ and $\text{inr}_{X,Y}: Y \rightarrow X + Y$ are the first and the second injections. Furthermore, the monoidal product distributes over the coproducts. Namely, the canonical s-finite kernel $\widehat{\text{dst}}_{X,Y,Z}: X \times Z + Y \times Z \rightsquigarrow (X + Y) \times Z$ given by

$$\text{dst}_{X,Y,Z}(\bullet, (x, z)) = ((\bullet, x), z), \quad \text{dst}_{X,Y,Z}(\circ, (y, z)) = ((\circ, y), z)$$

is a natural isomorphism.

Proof. For associativity of the composition, see [32, Lemma 3], and for functoriality of \otimes , see [32, Proposition 5]. It is not difficult to check that the category of measurable spaces and s-finite kernels associated with \otimes and 1 forms a symmetric monoidal category. For s-finite kernels $f: X \rightsquigarrow Z$ and $g: Y \rightsquigarrow Z$, the cotupling $[f, g]: X + Y \rightsquigarrow Z$ is given by

$$[f, g](\bullet, x) = f(x), \quad [f, g](\circ, y) = g(y).$$

It follows from universality of coproducts that $\widehat{\text{dst}}_{X,Y,Z}$ is a natural isomorphism. \square

For s-finite kernels $k: X \rightsquigarrow Y$ and $h: Z \rightsquigarrow W$, we define an s-finite kernel $k \oplus h: X + Z \rightsquigarrow Y + W$ by

$$(k \oplus h)(\bullet, x) = k(x) \text{ where } A_Y = \{y : (\bullet, y) \in A\}, \\ (k \oplus h)(\circ, z) = h(z) \text{ where } A_W = \{w : (\circ, w) \in A\}.$$

This is the unique s-finite kernel satisfying

$$(k \oplus h) \circ \text{inl}_{X,Z} = \text{inl}_{Y,W} \circ k, \quad (k \oplus h) \circ \text{inr}_{X,Z} = \text{inr}_{Y,W} \circ h.$$

10.2 Probabilistic Mealy Machine

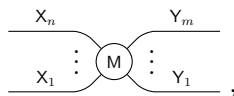
Definition 10.1. *For **Int**-objects X and Y , a probabilistic Mealy machine M from X to Y consists of*

- *a measurable space S_M called the state space of M ;*
- *an element $s_M \in S_M$ called the initial state of M ;*
- *an s-finite kernel $\tau_M: (X^+ + Y^-) \times S_M \rightsquigarrow (Y^+ + X^-) \times S_M$ called the transition relation.*

When M is a probabilistic Mealy machine from X to Y , we write $M: X \rightarrow Y$.

We can regard a Mealy machine $M: X \rightarrow Y$ as a probabilistic Mealy machine from X to Y by identifying the transition function $\tau_M: (X^+ + Y^-) \times S_M \rightarrow (Y^+ + X^-) \times S_M$ with the corresponding s-finite kernel $\widehat{\tau}_M: (X^+ + Y^-) \times S_M \rightsquigarrow (Y^+ + X^-) \times S_M$. In the sequel, we confuse Mealy machines (and token machines) with corresponding probabilistic Mealy machines.

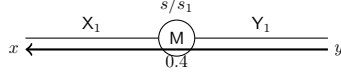
Let X_1, \dots, X_n and Y_1, \dots, Y_m be **Int**-object. Just like Mealy machines, we depict a probabilistic Mealy machine M from $X_1 \otimes \dots \otimes X_n$ to $Y_1 \otimes \dots \otimes Y_m$ as a box with edges labeled by X_1, \dots, X_n on the left hand side and edges labeled by Y_1, \dots, Y_m on the right hand side:



and we depict transitions as arrows. For example, when $n = m = 1$, we depict

$$\tau_M(((o, y), s), A) = 0.4[(((o, x), s_1) \in A)],$$

for $y \in Y_1^-$, $x \in X_1^-$ and $s, s_1, s_2 \in S_M$ as the following arrow



where the positive real on the arrow indicate probabilities of the transition. Below, we may omit states and probabilities of transitions when they are not important or are easy to infer.

10.3 Behavioral Equivalence

We give an equivalence relation between probabilistic Mealy machines so as to identify probabilistic Mealy machines that behaves in the same way. Let X and Y be **Int**-objects, and let M and N be probabilistic Mealy machines from X to Y . We write $M \sim_{X,Y} N$ when there is a measurable function $f: S_M \rightarrow S_N$ such that $f(s_M) = s_N$ and the following diagram commutes:

$$\begin{array}{ccc} (X^+ + Y^-) \times S_M & \xrightarrow{\text{id} \otimes \hat{f}} & (X^+ + Y^-) \times S_N \\ \tau_M \downarrow & & \downarrow \tau_N \\ (Y^+ + X^-) \times S_M & \xrightarrow{\text{id} \otimes \hat{f}} & (Y^+ + X^-) \times S_N \end{array}$$

We define an equivalence relation $\simeq_{X,Y}$ to be the symmetric transitive closure of $\sim_{X,Y}$. A probabilistic Mealy machine $M: X \rightarrow Y$ is *behaviorally equivalent* to $N: X \rightarrow Y$ when we have $M \simeq_{X,Y} N$. When we can infer subscripts of $\simeq_{X,Y}$, we omit them. We say that a measurable function $f: S_M \rightarrow S_N$ realizes a behavioral equivalence $M \simeq N$ (realizes $M \sim N$) when $M \sim N$ is witnessed by f .

10.4 Construction of probabilistic Mealy Machines

We introduce probabilistic Mealy machines and their constructions that are building blocks of our denotational semantics. Most of them are adaptation of Mealy machines in Section 5.4, and we just give their formal definitions.

10.4.1 Composition/Cut

For probabilistic Mealy machines $M: X \rightarrow Y$ and $N: Y \rightarrow Z$, we define the state space and the initial states of $N \circ M$ by $S_{N \circ M} = S_M \times S_N$, $s_{N \circ M} = (s_M, s_N)$ and we define the transition relation $\tau_{N \circ M}$ by

$$\tau_{N \circ M} = k_{X^+, Z^-, Z^+, X^-} \vee \bigvee_{n \in \mathbb{N}} k_{Y^+, Y^-, Z^+, X^-} \circ k_{Y^+, Y^-, Y^+, Y^-}^n \circ k_{X^+, Z^-, Y^+, Y^-}$$

where $k_{A,B,C,D}: (A + B) \times S_{N \circ M} \rightsquigarrow (C + D) \times S_{N \circ M}$ are restrictions of the following s-finite kernel

$$\begin{array}{c} (X^+ + Z^- + Y^+ + Y^-) \times S_{N \circ M} \\ \downarrow \cong \\ (X^+ + Y^-) \times S_M \times S_N + (Y^+ + Z^-) \times S_N \times S_M \\ \downarrow (\tau_M \otimes S_N) \oplus (\tau_N \otimes S_M) \\ (Y^+ + X^-) \times S_M \times S_N + (Z^+ + Y^-) \times S_N \times S_M \\ \downarrow \cong \\ (Z^+ + X^- + Y^+ + Y^-) \times S_{N \circ M} \end{array}$$

namely, the s-finite kernels $k_{A,B,C,D}$ satisfies

$$\begin{array}{ccc} A \times S_{\mathbf{N} \circ \mathbf{M}} & \xrightarrow{\text{inj, dst}} & ((X^+ + Y^-) \times S_{\mathbf{M}} \times S_{\mathbf{N}}) + ((Y^+ + Z^-) \times S_{\mathbf{N}} \times S_{\mathbf{M}}) \\ \downarrow k_{A,B} & & \downarrow (\tau_{\mathbf{M}} \otimes \text{id}_{S_{\mathbf{N}}}) \oplus (\tau_{\mathbf{N}} \otimes \text{id}_{S_{\mathbf{M}}}) \\ B \times S_{\mathbf{N} \circ \mathbf{M}} & \xrightarrow{\text{inj, dst}} & ((Y^+ + X^-) \times S_{\mathbf{M}} \times S_{\mathbf{N}}) + ((Z^+ + Y^-) \times S_{\mathbf{N}} \times S_{\mathbf{M}}) . \end{array}$$

Here, the horizontal arrows consists of the injection from A into $X^+ + Y^- + Y^+ + Z^-$ followed by distributivity and symmetry. For example, when $A = X^+ + Z^-$, the upper horizontal arrow is given by

$$\begin{array}{c} (X^+ + Z^-) \times S_{\mathbf{N} \circ \mathbf{M}} \\ \downarrow (\text{inl}_{X^+, Y^-} \oplus \text{inr}_{Y^+, Z^-}) \otimes \text{id}_{S_{\mathbf{N} \circ \mathbf{M}}} \\ ((X^+ + Y^-) + (Y^+ + Z^-)) \times S_{\mathbf{N} \circ \mathbf{M}} \\ \downarrow \text{dst}_{X^+ + Y^-, Y^+ + Z^-, S_{\mathbf{N} \circ \mathbf{M}}} \\ ((X^+ + Y^-) \times S_{\mathbf{N} \circ \mathbf{M}}) + ((Y^+ + Z^-) \times S_{\mathbf{N} \circ \mathbf{M}}) \\ \downarrow \text{id}_{(X^+ + Y^-) \times S_{\mathbf{N} \circ \mathbf{M}}} \oplus (\text{id}_{Y^+ + Z^-} \otimes \text{sym}_{S_{\mathbf{N}}, S_{\mathbf{M}}}) \\ ((X^+ + Y^-) \times S_{\mathbf{M}} \times S_{\mathbf{N}}) + ((Y^+ + Z^-) \times S_{\mathbf{N}} \times S_{\mathbf{M}}) . \end{array}$$

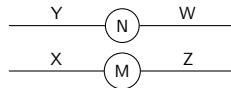
Joins in the definition of the composition of probabilistic Mealy machines are the pointwise ordoer. We can check that the composition of probabilistic Mealy machines is compatible with behavioural equivalence and that **Int**-objects and the composition of probabilistic Mealy machines is a category where the identity on an **Int**-object X is $\text{id}_X: X \rightarrow X$ (regarded as a probabilistic Mealy machine).

10.4.2 Monoidal Products

We give monoidal products of probabilistic Mealy machines. For probabilistic Mealy machines $M: X \rightarrow Z$ and $N: Y \rightarrow W$, we define a probabilistic Mealy machine $M \otimes N: X \otimes Y \rightarrow Z \otimes W$ by: $S_{M \otimes N} = S_M \times S_N$, $s_{M \otimes N} = (s_M, s_N)$ and $\tau_{M \otimes N}$ is given by

$$\begin{array}{c} ((X^+ + Y^+) + (W^- + Z^-)) \times S_{M \otimes N} \\ \downarrow \cong \\ (X^+ + Z^-) \times S_M \times S_N + (Y^+ + W^-) \times S_N \times S_M \\ \downarrow (\tau_M \otimes S_N) \oplus (\tau_N \otimes S_M) \\ (Z^+ + X^-) \times S_M \times S_N + (W^+ + Y^-) \times S_N \times S_M \\ \downarrow \cong \\ ((Z^+ + W^+) + (Y^- + X^-)) \times S_{M \otimes N} . \end{array}$$

It is not difficult to check that the monoidal product is compatible with behavioural equivalence. We depict $M \otimes N: (X \otimes Y) \rightarrow (Z \otimes W)$ as follows:



For unit_X , counit_X and $\text{sym}_{X,Y}$, we adopt the same diagrammatic presentation.

10.4.3 A Modal Operator

Let $M: X \rightarrow Y$ be a probabilistic Mealy machine. We define a probabilistic Mealy machine $!M: !X \rightarrow !Y$ by: the state space of $!M$ is defined to be $|M|^{\mathbb{N}}$ associated with the least σ -algebra

such that for all $A_1, A_2, \dots \in \Sigma_M$,

$$A_1 \times A_2 \times \dots \in \Sigma_M;$$

the initial state s_M is (s_M, s_M, \dots) ; the transition function τ_M is the unique partial measurable function satisfying

$$\begin{array}{ccc} (X^+ + Y^-) \times S_M \times S_M^{\mathbb{N}} & \xrightarrow{(\text{inj}_n \oplus \text{inj}_n) \otimes \text{ins}_n} & (\mathbb{N} \times X^+ + \mathbb{N} \times Y^-) \times S_M^{\mathbb{N}} \\ \downarrow \tau_M \otimes S_M^{\mathbb{N}} & & \downarrow \tau_M \\ (Y^+ + X^-) \times S_M \times S_M^{\mathbb{N}} & \xrightarrow{(\text{inj}_n \oplus \text{inj}_n) \otimes \text{ins}_n} & (\mathbb{N} \times Y^+ + \mathbb{N} \times X^-) \times S_M^{\mathbb{N}} \end{array}$$

for all $n \in \mathbb{N}$. Here, $\text{inj}_n: (-) \rightarrow \mathbb{N} \times (-)$ are the n th injections, and $\text{ins}_n: S_M \times S_M^{\mathbb{N}} \rightarrow S_M^{\mathbb{N}}$ sends $(s, \{s_n\}_{n \in \mathbb{N}})$ to $(s_0, \dots, s_{n-1}, s, s_n, s_{n+1}, \dots)$.

10.4.4 Diagrammatic Reasoning on Probabilistic Mealy Machines

Diagrammatic reasoning is valid also for probabilistic Mealy machines.

Proposition 10.2. *The category **pMealy** of **Int**-object and probabilistic Mealy machines (modulo behavioural equivalence) is a compact closed category. The dual of an **Int**-object X is X^\perp . The unit and the counit arrows are unit_X and counit_X .*

Proposition 10.3. *If two probabilistic Mealy machines have the same diagrammatic presentation modulo some rearrangement of edges and nodes, then they are behaviourally equivalent.*

We can check Proposition 10.2 by replacing the category of partial measurable functions by the category of s-finite kernels in Section 5.6.

10.4.5 A State Monad

We define an **Int**-object J by $(1, 1)$ and define an **Int**-object J_0 by $(1, \emptyset)$. Then $J \otimes (-)$ is a state monad (on **pMealy**), whose unit and multiplication are given by:

$$j \otimes X: X \rightarrow J \otimes X, \quad n \otimes X: J \otimes J \otimes X \rightarrow J \otimes X$$

where $j = \text{unit}_{J_0}$ and $n = J_0 \otimes \text{counit}_{J_0} \otimes J_0^\perp$.

10.4.6 Scoring

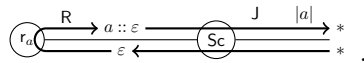
We construct a probabilistic Mealy machine $\text{Sc}: R \rightarrow J$ by:

$$S_{\text{Sc}} = 1, \quad s_{\text{Sc}} = *$$

and

$$\tau_{\text{Sc}}(((\bullet, u), *), A) = \begin{cases} |a| [((\bullet, *), *) \in A], & \text{if } u = a :: v, \\ 0, & \text{otherwise,} \end{cases} \quad \tau_{\text{Sc}}(((\circ, *), *), A) = [((\circ, \varepsilon), *) \in A].$$

The probabilistic Mealy machine simulates scoring $\text{score}(r_a)$ as follows:



10.4.7 Sampling

We define a Mealy machine $\text{Sa}: \mathbf{I} \rightarrow \mathbf{J} \otimes !\mathbf{R}$ by: the state space S_{Sa} is defined to be $\{*\} \cup \mathbb{R}_{[0,1]}$, and the initial state s_{Sa} is $*$, and the transition function

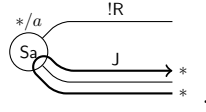
$$\tau_{\text{Sa}}: (\emptyset + (\mathbb{N} \times \mathbb{S} + 1)) \times S_{\text{Sa}} \rightsquigarrow ((1 + \mathbb{N} \times \mathbb{S}) + \emptyset) \times S_{\text{Sa}}$$

is given by

$$\begin{aligned} \tau_{\text{Sa}}(((\circ, (\bullet, (n, u))), s), A) &= [((\bullet, (\circ, (n, s :: u))), s) \in A] [s \in \mathbb{R}], \\ \tau_{\text{Sa}}(((\circ, (\circ, *)), s), A) &= [s = *] \mu_{\text{Borel}}(\{a \in \mathbb{R}_{[0,1]} : ((\bullet, (\bullet, *)), a) \in A\}). \end{aligned}$$

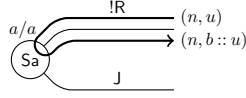
The probabilistic Mealy machine behaves as follows:

- In the initial state $*$, given $*$ from the \mathbf{J} -edge, Sa draws a real number from the uniform distribution and stores the real number:



For example, the probability of the state being a real number in $[0, 0.3]$ after this transition is 0.3.

- After this transition, Sa returns $(n, a :: u)$ to each “query” (n, u) :



11 Probabilistic Mealy Machine Semantics for PCFSS

We interpret a type \mathbf{A} as the **Int**-object $\llbracket \mathbf{A} \rrbracket_{\mathbf{d}}$ given by

$$\llbracket \mathbf{Unit} \rrbracket_{\mathbf{d}} = \mathbf{I}, \quad \llbracket \mathbf{Real} \rrbracket_{\mathbf{d}} = \mathbf{R}, \quad \llbracket \mathbf{A} \rightarrow \mathbf{B} \rrbracket_{\mathbf{d}} = \mathbf{J} \otimes !\llbracket \mathbf{B} \rrbracket_{\mathbf{d}} \otimes !\llbracket \mathbf{A} \rrbracket_{\mathbf{d}}^{\perp}.$$

We define interpretation of contexts by

$$\llbracket \mathbf{x} : \mathbf{A}, \dots, \mathbf{y} : \mathbf{B} \rrbracket_{\mathbf{d}} = \llbracket \mathbf{A} \rrbracket_{\mathbf{d}} \otimes \dots \otimes \llbracket \mathbf{B} \rrbracket_{\mathbf{d}}.$$

When Δ is the empty sequence, we define $\llbracket \Delta \rrbracket_{\mathbf{d}}$ to be \mathbf{I} .

We interpret terms $\Delta \vdash \mathbf{M} : \mathbf{A}$ and values $\Delta \vdash \mathbf{V} : \mathbf{A}$ by

$$\llbracket \Delta \vdash \mathbf{M} : \mathbf{A} \rrbracket_{\mathbf{d}} : !\llbracket \Delta \rrbracket_{\mathbf{d}} \rightarrow \mathbf{S} \otimes !\llbracket \mathbf{A} \rrbracket_{\mathbf{d}}, \quad (\Delta \vdash \mathbf{V} : \mathbf{A})_{\mathbf{d}} : !\llbracket \Delta \rrbracket_{\mathbf{d}} \rightarrow \llbracket \mathbf{A} \rrbracket_{\mathbf{d}}$$

inductively defined by diagrams in Figure 6 where Mealy machines are regarded as probabilistic Mealy machines in the obvious manner.

11.1 Soundness and Adequacy

11.1.1 Observation

Let $\mathbf{M}: \mathbf{I} \rightarrow \mathbf{J} \otimes !\mathbf{R}$ be a probabilistic Mealy machine. We define s-finite kernels $t_0: \mathbf{1} \rightsquigarrow S_{\mathbf{M}}$ and $t_1: S_{\mathbf{M}} \rightsquigarrow \mathbb{R}$ by

$$\begin{aligned} t_0^{\mathbf{M}}(*, A) &= \tau_{\mathbf{M}}(((\circ, (\circ, *)), s_{\mathbf{M}}), \{((\bullet, (\bullet, *)), s) : s \in A\}), \\ t_1^{\mathbf{M}}(s, A) &= \tau_{\mathbf{M}}(((\circ, (\bullet, (0, \varepsilon))), s), \{(\bullet, (\circ, (0, a :: \varepsilon))) : a \in A\}). \end{aligned}$$

Then we define a measure $\text{obs}(\mathbf{M})$ on \mathbb{R} to be $t_1^{\mathbf{M}} \circ t_0^{\mathbf{M}}(*, -)$. Intuitively, $\text{obs}(\mathbf{M})$ is a measure that describes distribution of real numbers obtained by the following process:

Lemma 11.2. *For all closed terms $M, N : \mathbf{A}$, if $M \xrightarrow{\text{red}} N$, then $\llbracket M \rrbracket = \llbracket N \rrbracket$.*

Proof. By case analysis. For the case of recursion, see Corollary 11.1. □

We first prove soundness.

Proposition 11.1. *For any closed term $M : \mathbf{Real}$, if $M \Rightarrow_n \mu$, then $\mu \leq \text{obs}(\llbracket M \rrbracket_d)$.*

Proof. By induction on n . (Base case) Easy. (Induction step) By case analysis.

- If $M = r_a$, then $\text{obs}(\llbracket M \rrbracket_d) = \delta_a$.
- If $M = E[N]$ and $N \xrightarrow{\text{red}} L$, then $\text{obs}(\llbracket M \rrbracket_d) = \text{obs}(\llbracket E[L] \rrbracket_d) \geq \mu$.
- If $M = E[\text{score}(r_a)]$ and $E[\text{skip}] \Rightarrow_{n-1} \mu$, then by the definition of $\text{obs}(-)$, we see that $\text{obs}(\llbracket M \rrbracket_d) = |a| \text{obs}(\llbracket E[\text{skip}] \rrbracket_d) \geq |a| \mu$.
- If $M = E[\text{sample}]$ and $E[r_a] \Rightarrow_{n-1} k(a, -)$ for some finite kernel k , then by the definition of t_0 and t_1 , we see that

$$\begin{aligned} t_0^{\llbracket M \rrbracket_d} &= 1 \rightsquigarrow^{\mu_{\text{Borel}}} \mathbb{R} \rightsquigarrow^h S_{\llbracket M \rrbracket_d} \times \mathbb{R} \rightsquigarrow^{\text{injection}} S_{\llbracket M \rrbracket_d} \times (1 + \mathbb{R}), \\ t_1^{\llbracket M \rrbracket_d}((s, a), -) &= t_1^{\llbracket E[r_a] \rrbracket_d}(s, -), \\ t_1^{\llbracket M \rrbracket_d}((s, 0), -) &= 0 \end{aligned}$$

for some h such that $h(a, -) = t_0^{\llbracket E[r_a] \rrbracket_d} \times \delta_a$. Hence,

$$(t_1^{\llbracket M \rrbracket_d} \circ t_0^{\llbracket M \rrbracket_d})(*, A) = \int_{\mathbb{R}_{[0,1]}} (t_1^{\llbracket E[r_a] \rrbracket_d} \circ t_0^{\llbracket E[r_a] \rrbracket_d})(*, A) da \geq \int_{\mathbb{R}_{[0,1]}} k(a, A) da.$$

□

It remains to prove that $M \Rightarrow_\infty \mu$ implies $\text{obs}(\llbracket M \rrbracket_d) \leq \mu$. We use logical relations. We define a binary relation O_d between closed terms of type \mathbf{Real} and probabilistic Mealy machines from I to $J \otimes !R$ by

$$(M, M) \in O_d \iff \text{if } M \Rightarrow_\infty \mu \text{ then } \text{obs}(M) \leq \mu.$$

We then inductively define binary relations

$$\begin{aligned} S_A &\subseteq \{\text{closed values of type } \mathbf{A}\} \times \{\text{Mealy machines from } I \text{ to } \llbracket \mathbf{A} \rrbracket\} \\ S_A^\top &\subseteq \{\text{evaluation contexts } \mathbf{x} : \mathbf{A} \vdash E[\mathbf{x}] : \mathbf{Real}\} \times \{\text{Mealy machines from } \llbracket \mathbf{A} \rrbracket \text{ to } J \otimes !R\} \\ \overline{S}_A &\subseteq \{\text{closed terms of type } \mathbf{A}\} \times \{\text{Mealy machines from } I \text{ to } J \otimes \llbracket \mathbf{A} \rrbracket\} \end{aligned}$$

by

$$\begin{aligned} S_{\mathbf{Real}} &= \{(r_a, r_a) : a \in \mathbb{R}\}, \\ S_{\mathbf{Unit}} &= \{(\text{skip}, \text{id}_I)\}, \\ S_{A \rightarrow B} &= \{(V, M) : \forall (W, N) \in S_A, (VW, (J \otimes \llbracket B \rrbracket_d \otimes \text{counit}_{\llbracket A \rrbracket_d}) \circ (M \otimes !N)) \in \overline{S}_B\}, \\ S_A^\top &= \{(E[-], E) : \forall (V, M) \in S_A, (E[V], E \circ !M) \in O_d\}, \\ \overline{S}_A &= \{(M, M) : \forall (E[-], E) \in S_A^\top, (E[M], (n \otimes !R) \circ (J \otimes E) \circ M) \in O_d\} \end{aligned}$$

We list some properties of the logical relations.

Lemma 11.3. *Let \mathbf{A} be a type.*

1. *If $(V, M) \in S_A$, then $(V, j \otimes !M) \in \overline{S}_A$.*

2. If $(M, M) \in \overline{S}_A$ and $N \xrightarrow{\text{red}} M$, then $(N, M) \in \overline{S}_A$.
3. If $(M, M) \in \overline{S}_A$ and $M \xrightarrow{\text{red}} N$, then $(N, M) \in \overline{S}_A$.
4. If $(M, M) \in \overline{S}_A$ and $M \simeq N$, then $(M, N) \in \overline{S}_A$.
5. For any closed term $M : A$, $(M, \text{bot}_{J \otimes ![[A]]}) \in \overline{S}_A$ where $\text{bot}_X : I \rightarrow X$ is a token machine whose transition function is the zero kernel.
6. For any closed value $V : A \rightarrow B$, $(V, \text{bot}_{[[A \rightarrow B]]}) \in S_{A \rightarrow B}$.
7. If $(M, M_i) \in \overline{S}_A$ and $S_{M_1} = S_{M_2} = \dots$ and $s_{M_1} = s_{M_2} = \dots$ and $\tau_{M_1} \leq \tau_{M_2} \leq \dots$, then $(M, N) \in \overline{S}_A$ where N is given by $S_N = S_{M_1}$, $s_N = s_{M_1}$, $\tau_N = \bigvee_n \tau_{M_n}$.

Proof. We can check these items by unfolding the definition of O_d and the logical relations. \square

Lemma 11.4 (Basic Lemma). *Let $\Delta = (x : A_1, \dots, x_n : A_n)$ be a context.*

- For any term $\Delta \vdash M : A$ and for any $(V_i, N_i) \in S_{A_i}$ for $i = 1, 2, \dots, n$, we have

$$(M\{V_1/x_1, \dots, V_n/x_n\}, [[M]]_d \circ (!N_1 \otimes \dots \otimes !N_n)) \in \overline{S}_A.$$

- For any value $\Delta \vdash V : A$ and for any $(V_i, N_i) \in S_{A_i}$ for $i = 1, 2, \dots, n$, we have

$$(V\{V_1/x_1, \dots, V_n/x_n\}, (M)_d \circ (!N_1 \otimes \dots \otimes !N_n)) \in S_A.$$

Proof. By induction on M and V . Most cases follow from Lemma 11.3. For $M = \text{sample}$ and $M = \text{score}(V)$, we check the statement by unfolding the definition of S_a and s_c . Here, we only check for $M = \text{sample}$ and $M = \text{fix}_{A,B}(f, x, N)$.

- When $M = \text{sample}$, let (E, E) be a pair in R_{Real}^\top . We write N for $(n \otimes !R) \circ (J \otimes E) \circ S_a$. By the definition of t_0 and t_1 , we have

$$\begin{aligned} t_0^N &= 1 \xrightarrow{\mu_{\text{Borel}}} \mathbb{R} \xrightarrow{h} S_N \times \mathbb{R} \xrightarrow{\text{injection}} S_N \times (1 + \mathbb{R}), \\ t_1^N((s, a), -) &= t_1^{\text{Eora}}(s, -), \\ t_1^N((s, 0), -) &= 0 \end{aligned}$$

for some h such that $h(a, -) = t_0^{\text{Eora}} \times \delta_a$. Hence,

$$(t_1^N \circ t_0^N)(*, A) = \int_{\mathbb{R}_{[0,1]}} (t_1^{\text{Eora}} \circ t_0^{\text{Eora}})(*, A) \, da = \int_{\mathbb{R}_{[0,1]}} k(a, A) \, da.$$

- When $M = \text{fix}_{A,B}(f, x, N)$, for simplicity, we suppose that M is a closed term. By induction hypothesis, we can check that

$$(M, (\lambda x^A. N)_d \circ !(\lambda x^A. N)_d \circ \dots \circ !^k(\lambda x^A. N)_d \circ \text{bot}_{I, !^k[[A \rightarrow B]]_d}) \in R_{A \rightarrow B}$$

by induction on n . By Lemma 11.3 and Proposition 11.2, we obtain $(M, [[M]]) \in R_{A \rightarrow B}$. \square

Theorem 11.2. *For any closed term $\vdash M : \text{Real}$, if $M \Rightarrow_\infty \mu$, then*

$$\text{obs}([M]_d) \leq \mu.$$

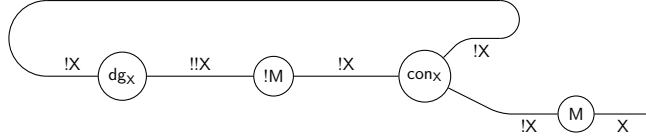
Proof. By soundness, we have $\mu \leq \text{obs}([M]_d)$. On the other hand, because $([-], j \otimes \text{id}_{!R})$ is an element of S_{Real}^\top , we obtain the other inequality by Lemma 11.4. \square

11.1.3 Induction step on recursion

Our Goal: Approximation Lemma Let $M : !X \rightarrow X$ be a Mealy machine. In this section, we show that a Mealy machine $M^\dagger : I \rightarrow !X$ given by

$$M^\dagger = (\text{counit}_{!X} \otimes M) \circ ((\text{con}_X \circ !M \circ \text{dg}_X) \otimes \text{id}_{!X}) \circ \text{unit}_{!X}.$$

is a “least” fixed point of M . Diagrammatically, M^\dagger consists of digging, contraction and a feed back loop:



This construction already appeared in the interpretation of the fixed point operator. In fact, for a term $f : A \rightarrow B, x : A \vdash M : B$, we have $\llbracket \text{fix}_{A,B}(f, x, M) \rrbracket = \llbracket \lambda x^A. M \rrbracket^\dagger$.

Parametrized Modal Operator and Parametrized Loop Operator We introduce parametrization of the modal operator $!$ and the loop operator $(-)^{\dagger}$. For $\alpha \subseteq \mathbb{N}$, we define $!_{\alpha}M$ by: the state space and the initial state of $!M$ are given by

$$S_{!_{\alpha}M} = S_{!M} = S_M^{\mathbb{N}}, \quad s_{!_{\alpha}M} = s_{!M}$$

and $\tau_{!_{\alpha}M}$ is a unique s-finite kernel such that the following diagrams commute:

- for any $n \in \alpha$,

$$\begin{array}{ccc} (X^+ + Y^-) \times S_M \times S_{!_{\alpha}M} & \xrightarrow{(\text{inj}_n \oplus \text{inj}_n) \otimes \text{ins}_n} & ((\mathbb{N} \times X^+) + (\mathbb{N} \times Y^-)) \times S_{!_{\alpha}M} \\ \tau_M \otimes S_{!_{\alpha}M} \Big\downarrow & & \Big\downarrow \tau_{!_{\alpha}M} \\ (Y^+ + X^-) \times S_M \times S_{!_{\alpha}M} & \xrightarrow{(\text{inj}_n \oplus \text{inj}_n) \otimes \text{ins}_n} & ((\mathbb{N} \times Y^+) + (\mathbb{N} \times X^-)) \times S_{!_{\alpha}M} \end{array}$$

- for any $n \notin \alpha$,

$$\begin{array}{ccc} (X^+ + Y^-) \times S_M \times S_{!_{\alpha}M} & \xrightarrow{(\text{inj}_n \oplus \text{inj}_n) \otimes \text{ins}_n} & ((\mathbb{N} \times X^+) + (\mathbb{N} \times Y^-)) \times S_{!_{\alpha}M} \\ \tau_M \otimes S_{!_{\alpha}M} \Big\downarrow & & \Big\downarrow \emptyset \\ (Y^+ + X^-) \times S_M \times S_{!_{\alpha}M} & \xrightarrow{(\text{inj}_n \oplus \text{inj}_n) \otimes \text{ins}_n} & ((\mathbb{N} \times Y^+) + (\mathbb{N} \times X^-)) \times S_{!_{\alpha}M} \end{array}$$

Let $\alpha_n, \beta_n \subseteq \mathbb{N}$ be

$$\alpha_0 = \emptyset, \quad \beta_n = \{\langle i, j \rangle : i \in \alpha_n \text{ and } j \in \mathbb{N}\}, \quad \alpha_{n+1} = \{2i : i \in \mathbb{N}\} \cup \{2i + 1 : i \in \beta_n\}.$$

The definition of α_n and β_n are motivated by the following lemma.

Lemma 11.5. For any $n \in \mathbb{N}$ and for any $M : X \rightarrow Y$, we have

$$\text{con}_Y \circ !_{\alpha_{n+1}} M \simeq (!M \otimes !_{\beta_n} M) \circ \text{con}_X, \quad \text{dg}_Y \circ !_{\beta_{n+1}} M \simeq !_{\alpha_n} !M \circ \text{dg}_X.$$

By means of $!_{\alpha}$, we also parametrize the operator $(-)^{\dagger}$. For $\alpha \subseteq \mathbb{N}$, and for $M : !X \rightarrow X$, we define $M^{\dagger, \alpha} : I \rightarrow !X$ by

$$M^{\dagger, \alpha} = (\text{counit}_{!X} \otimes M) \circ ((\text{con}_X \circ !_{\alpha} M \circ \text{dg}_X) \otimes \text{id}_{!X}) \circ \text{unit}_{!X}.$$

Because $!_{\mathbb{N}} M = !M$, we have $M^\dagger = M^{\dagger, \mathbb{N}}$.

Lemma 11.6. For any $\alpha \subseteq \mathbb{N}$, we have

$$S_{M^\dagger, \alpha} = S_{M^\dagger}, \quad s_{M^\dagger, \alpha} = s_{M^\dagger}.$$

Below, we write $h(M): S_{M^\dagger} \rightarrow S_M \times S_M^{\mathbb{N}}$ for the isomorphism obtained by applying $1 \times (-) \cong (-)$ to S_M .

Lemma 11.7. There is a family of measurable functions $\phi_X: (X^{\mathbb{N}})^{\mathbb{N}} \rightarrow (X^{\mathbb{N}})^{\mathbb{N}}$ such that the following diagram commutes:

$$\begin{array}{ccc} (X^{\mathbb{N}})^{\mathbb{N}} & & \\ \downarrow u_{X^{\mathbb{N}}} & \searrow \phi_X & \\ (X^{\mathbb{N}})^{\mathbb{N}} \times ((X^{\mathbb{N}})^{\mathbb{N}})^{\mathbb{N}} & & \\ \downarrow (X^{\mathbb{N}})^{\mathbb{N}} \times \phi_{X^{\mathbb{N}}} & & \\ (X^{\mathbb{N}})^{\mathbb{N}} \times ((X^{\mathbb{N}})^{\mathbb{N}})^{\mathbb{N}} & \xrightarrow{\cong} & (X^{\mathbb{N}} \times (X^{\mathbb{N}})^{\mathbb{N}})^{\mathbb{N}} \xrightarrow{(u_X^{-1})^{\mathbb{N}}} (X^{\mathbb{N}})^{\mathbb{N}}. \end{array}$$

where $u_X: X^{\mathbb{N}} \rightarrow X^{\mathbb{N}} \times (X^{\mathbb{N}})^{\mathbb{N}}$ is a measurable isomorphism given by

$$u_X(x_n)_{n \in \mathbb{N}} = ((x_{2n})_{n \in \mathbb{N}}, ((x_{2\langle m_0, m_1 \rangle + 1})_{m_0 \in \mathbb{N}})_{m_1 \in \mathbb{N}}).$$

Proof. In this proof, for sets N_1, N_2, \dots, N_k , we identify elements in $((X^{N_1})^{N_2}) \dots^{N_k}$ with functions from $N_1 \times N_2 \times \dots \times N_k$ to X . For $x \in (X^{\mathbb{N}})^{\mathbb{N}}$ and $(a, b) \in \mathbb{N}$, We define $(\phi_X(x))(a, b)$ by induction on a :

$$(\phi_X(x))(a, b) = \begin{cases} x(a', 2b), & \text{if } a = 2a'; \\ (\phi_{X^{\mathbb{N}}}(x'))(a_0, a_1, b), & \text{if } a = 2\langle a_0, a_1 \rangle + 1. \end{cases}$$

where $x' \in ((X^{\mathbb{N}})^{\mathbb{N}})^{\mathbb{N}}$ is given by

$$x'(n, m, l) = x(n, 2\langle m, l \rangle + 1).$$

We note that this definition makes sense because $a = 2\langle a_0, a_1 \rangle + 1$ implies $a_1 < a$. It is straightforward to check the family ϕ_X makes the above diagram commute. \square

Lemma 11.8. For any Mealy machine $M: !X \rightarrow X$, we have

$$M^{\dagger, \alpha_{n+1}} \simeq M \circ (!M)^{\dagger, \alpha_n},$$

which is realized by u'_M given by

$$S_{M^{\dagger, \alpha_{n+1}}} \cong S_M \times S_M^{\mathbb{N}} \xrightarrow{S_M \times u_{S_M}} S_M \times S_M^{\mathbb{N}} \times (S_M^{\mathbb{N}})^{\mathbb{N}} \cong S_{M \circ (!M)^{\dagger, \alpha_n}}$$

where the first and the last isomorphisms are obtained by applying canonical isomorphisms $1 \times (-) \cong (-)$ and $(-) \times 1 \cong (-)$. \square

Proof. See Figure 5.

Lemma 11.9. For any Mealy machine $M: !X \rightarrow X$ and for any $n \in \mathbb{N}$,

$$(!M)^{\dagger, \alpha_n} \simeq !(M^{\dagger, \alpha_n})$$

which is realized by ϕ'_M given by

$$S_{(!M)^{\dagger, \alpha_n}} \cong S_M^{\mathbb{N}} \times (S_M^{\mathbb{N}})^{\mathbb{N}} \xrightarrow{S_M^{\mathbb{N}} \times \phi_{S_M}} S_M^{\mathbb{N}} \times (S_M^{\mathbb{N}})^{\mathbb{N}} \xrightarrow{\cong} (S_M \times S_M^{\mathbb{N}})^{\mathbb{N}} \cong S_{!(M^{\dagger, \alpha_n})}$$

where the first and the last isomorphisms are obtained by applying canonical isomorphisms $1 \times (-) \cong (-)$ and $(-) \times 1 \cong (-)$.

Proof. We prove the statement by induction on n . The base case follows from that the transition relations of $(!M)^{\dagger, \alpha_0}$ and $!(M^{\dagger, \alpha_0})$ are the zero kernels. We next check the induction step. We have

$$\begin{aligned}
(!M)^{\dagger, \alpha_{n+1}} &\simeq !M \circ (!M)^{\dagger, \alpha_n} && \text{(Lemma 11.8)} \\
&\simeq !M \circ !(M)^{\dagger, \alpha_n} && \text{(Induction hypothesis)} \\
&\simeq !(M \circ (!M)^{\dagger, \alpha_n}) && \text{(Proposition 5.3)} \\
&\simeq !(M^{\dagger, \alpha_{n+1}}). && \text{(Lemma 11.8)}
\end{aligned}$$

By Lemma 11.7, this behavioral equivalence is realized by ϕ'_M . \square

Proposition 11.2. *For a Mealy machine $M: !X \rightarrow X$, we inductively define $\text{iter}_n(M): ! \rightarrow X$ by*

$$\text{iter}_0(M) = \text{bot}_{!, X}, \quad \text{iter}_{n+1}(M) = M \circ !(\text{iter}_n(M)).$$

For all $n \in \mathbb{N}$, we have

$$S_{M^\dagger} = S_{M^\dagger, \alpha_n}, \quad s_{M^\dagger} = s_{M^\dagger, \alpha_n}, \quad M^{\dagger, \alpha_n} \simeq \text{iter}_n(M),$$

and

$$\tau_{M^\dagger, \alpha_0} \leq \tau_{M^\dagger, \alpha_1} \leq \tau_{M^\dagger, \alpha_2} \leq \cdots, \quad \tau_{M^\dagger} = \bigvee_{n \geq 0} \tau_{M^\dagger, \alpha_n}.$$

Proof. It follows from the definition of $!_{\alpha_n}$, we have

$$S_{!M} = S_{!_{\alpha_n} M}, \quad s_{!M} = s_{!_{\alpha_n} M}$$

for all $n \in \mathbb{N}$, and

$$\tau_{!_{\alpha_0} M} \leq \tau_{!_{\alpha_1} M} \leq \tau_{!_{\alpha_2} M} \leq \cdots, \quad \tau_{!M} = \bigvee_{n \geq 0} \tau_{!_{\alpha_n} M}$$

Hence, by continuity of the composition, the coproduct and the monoidal product of s-finite kernels, we have

$$S_{M^\dagger} = S_{M^\dagger, \alpha_n}, \quad s_{M^\dagger} = s_{M^\dagger, \alpha_n}$$

for all $n \in \mathbb{N}$, and

$$\tau_{M^\dagger, \alpha_0} \leq \tau_{M^\dagger, \alpha_1} \leq \tau_{M^\dagger, \alpha_2} \leq \cdots, \quad \tau_{M^\dagger} = \bigvee_{n \geq 0} \tau_{M^\dagger, \alpha_n}.$$

By induction on $n \in \mathbb{N}$, we show that we have

$$\text{iter}_n(M) \simeq M^{\dagger, \alpha_n},$$

which is realized by f_n . For the base case, we have $M^{\dagger, \emptyset} \simeq \text{iter}_0(M)$ because these Mealy machines $M^{\dagger, \emptyset}$ and $\text{iter}_0(M)$ are behaviorally equivalent to $\text{emp}_{!, X}$. The induction step follows from Lemma 11.8. \square

Corollary 11.1. *For any Mealy machine $M: !X \rightarrow X$,*

$$M \circ !M^\dagger \simeq M^\dagger.$$

Proof. By Proposition 9.1, we have

$$S_{M \circ !M^\dagger} = S_{M \circ !(M^\dagger, \alpha_n)}, \quad s_{M \circ !M^\dagger} = s_{M \circ !(M^\dagger, \alpha_n)},$$

and

$$\tau_{M \circ !M^\dagger} = \bigvee_{n \in \mathbb{N}} \tau_{M \circ !(M^\dagger, \alpha_n)}.$$

Therefore, by Lemma 11.8 and Lemma 11.9, the following diagram commutes:

$$\begin{array}{ccc}
X^- \times S_M^{\mathbb{N}} & \xrightarrow{X^+ \otimes (u'_M \circ (S_M \times \phi'_M))^{\wedge}} & X^- \times S_M^{\mathbb{N}} \\
\downarrow \bigvee_{n \in \mathbb{N}} \tau_{M \circ !(\mathbf{M}^\dagger, \alpha_{n+1})} & & \downarrow \bigvee_{n \in \mathbb{N}} \tau_{\mathbf{M}^\dagger, \alpha_{n+1}} \\
X^- \times S_M^{\mathbb{N}} & \xrightarrow{X^+ \otimes (u'_M \circ (S_M \times \phi'_M))^{\wedge}} & X^- \times S_M^{\mathbb{N}}.
\end{array}$$

It is easy to see that $u'_M \circ (S_M \times \phi'_M)$ preserves the initial states. Hence, $M \circ !(\mathbf{M}^\dagger) \simeq \mathbf{M}^\dagger$. \square

11.1.4 Commutativity Modulo Observational Equivalence

Definition 11.1. For terms $\vdash M, N : A$, we say that M is observationally equivalent to N when for all context $C[-]$, if $C[M] \Rightarrow_\infty \mu$, then $C[N] \Rightarrow_\infty \mu$.

In this section, as an application of our GoI semantics, we show that for all $\vdash M : A$, $\vdash N : B$ and $x : A, y : B \vdash L : C$,

$$\text{let } x \text{ be } M \text{ in let } y \text{ be } N \text{ in } L$$

is observationally equivalent to

$$\text{let } y \text{ be } N \text{ in let } x \text{ be } M \text{ in } L.$$

To prove this equivalence, let O'_d be a binary relation between closed terms of type **Real** and probabilistic Mealy machines from I to $J \otimes !R$ by

$$(M, M) \in O'_d \iff (M, M) \in O_d \text{ and (Condition1) and (Condition2)}$$

where (Condition 1) is: for any $A \in \Sigma_{((1+\mathbb{N} \times \mathbb{S}) + \emptyset) \times S_M}$ such that

$$A = \{((\bullet, (\circ, (n, u))), s) \mid ((\bullet, (\circ, (n, u))), s) \in A\},$$

and for any $s \in S_M$, $\tau_M(((\circ, (\circ, *)), s), A) = 0$; (Condition 2) is: for any $A \in \Sigma_{((1+\mathbb{N} \times \mathbb{S}) + \emptyset) \times S_M}$ such that

$$A = \{((\bullet, (\bullet, *)), s) \mid ((\bullet, (\bullet, *)), s) \in A\},$$

and for any $(n, u) \in \mathbb{N} \times \mathbb{S}$, for any $s \in S_M$, $\tau_M(((\circ, (\bullet, (n, u))), s), A) = 0$. We then inductively define binary relations

$$\begin{aligned}
T_A &\subseteq \{\text{closed values of type } A\} \times \{\text{Mealy machines from } I \text{ to } \llbracket A \rrbracket\} \\
T_A^\top &\subseteq \{\text{evaluation contexts } x : A \vdash E[x] : \text{Real}\} \times \{\text{Mealy machines from } \llbracket A \rrbracket \text{ to } J \otimes !R\} \\
\bar{T}_A &\subseteq \{\text{closed terms of type } A\} \times \{\text{Mealy machines from } I \text{ to } J \otimes \llbracket A \rrbracket\}
\end{aligned}$$

by replacing O_d in the definition of S_A , S_A^\top and \bar{S}_A with O'_d . Then we can prove basic lemma for this logical relation.

Lemma 11.10 (Basic Lemma). Let $\Delta = (x : A_1, \dots, x_n : A_n)$ be a context.

- For any term $\Delta \vdash M : A$ and for any $(V_i, N_i) \in T_{A_i}$ for $i = 1, 2, \dots, n$, we have

$$(M\{V_1/x_1, \dots, V_n/x_n\}, \llbracket M \rrbracket_d \circ (!N_1 \otimes \dots \otimes !N_n)) \in \bar{T}_A.$$

- For any value $\Delta \vdash V : A$ and for any $(V_i, N_i) \in T_{A_i}$ for $i = 1, 2, \dots, n$, we have

$$(V\{V_1/x_1, \dots, V_n/x_n\}, \llbracket V \rrbracket_d \circ (!N_1 \otimes \dots \otimes !N_n)) \in T_A.$$

Proof. Almost equivalent to the proof of Lemma 11.4. \square

Corollary 11.2. *For any $\vdash M : A$,*

- *for any $A \in \Sigma_{((1+\mathbb{N} \times \llbracket A \rrbracket_d^+) + \emptyset) \times S_M}$ such that*

$$A = \{((\bullet, (\circ, (n, a))), s) \mid ((\bullet, (\circ, (n, a))), s) \in A\},$$

*and for any $s \in S_M$, $\tau_M(((\circ, (\circ, *)), s), A) = 0$;*

- *for any $A \in \Sigma_{((1+\mathbb{N} \times \llbracket A \rrbracket_d^+) + \emptyset) \times S_M}$ such that*

$$A = \{((\bullet, (\bullet, *)), s) \mid ((\bullet, (\bullet, *)), s) \in A\},$$

and for any $(n, a) \in \mathbb{N} \times \llbracket A \rrbracket_d^-$, for any $s \in S_M$, $\tau_M(((\circ, (\bullet, (n, a))), s), A) = 0$.

By Corollary 11.2 and by the definition of composition of probabilistic Mealy machines, we see that if

$$S_1 \otimes S_2 \xrightarrow{k \otimes S_2} S_1 \otimes S_2 \xrightarrow{\cong} S_2 \otimes S_1 \xrightarrow{h \otimes S_1} S_2 \otimes S_1 \xrightarrow{\cong} S_1 \otimes S_2 = S_1 \otimes S_2 \xrightarrow{k \otimes h} S_1 \otimes S_2 \quad (1)$$

then

$$\llbracket \text{let } x \text{ be } M \text{ in let } y \text{ be } N \text{ in } L \rrbracket_d = \llbracket \text{let } y \text{ be } N \text{ in let } x \text{ be } M \text{ in } L \rrbracket_d \quad (2)$$

where k and h are s-finite kernels given by restricting the domain and the codomain of $\tau_{\llbracket M \rrbracket_d}$ and $\tau_{\llbracket N \rrbracket_d}$ respectively. Because of commutativity for s-finite kernels [32], the equality (1) is true. Hence, (2) holds. Then by adequacy, we see that

$$\text{let } x \text{ be } M \text{ in let } y \text{ be } N \text{ in } L$$

is observationally equivalent to

$$\text{let } y \text{ be } N \text{ in let } x \text{ be } M \text{ in } L.$$

12 Conclusion

We introduced a denotational semantics for **PCFSS**, a higher-order functional language with sampling from a uniform continuous distribution and scoring. Following [28], we considered two operational semantics, namely a *distribution-based* operational semantics, which associates terms with distributions over real numbers, and a *sampling-based* operational semantics, which associates each term with a weight along every probabilistic branch. Our main results are adequacy theorems for both kinds of operational semantics, and it follows from these theorems that sampling-based operational semantics is essentially equivalent to distribution-based operational semantics. Another consequence of adequacy theorems is the possibility of diagrammatic reasoning for observational equivalence of programs. It follows from the observation in Section 5.5 and the adequacy theorems, that diagrammatic equivalence for denotation of terms implies observational equivalence. It would be interesting to explore possible connections between our work and other works on diagrammatic reasoning for probabilistic computation, such as [48, 49].

At this point, our language does not support any normalisation mechanism as a first class operator. However, capturing inference algorithms such as the Metropolis-Hastings algorithm [50, 51], which is structured around a number of interactions between programs and their environment, seems plausible. Exploring the relationships between “idealised” normalisation mechanisms and such “approximating” normalisation mechanisms from the point of view of GoI is an interesting topic for future work.

Acknowledgment

The authors are partially supported by the INRIA/JSPS project “CRECOGI”, and would like to thank Michele Pagani for many fruitful discussions about an earlier version of this work. Naohiko Hoshino is supported by JST ERATO HASUO Metamathematics for Systems Design Project (No. JPMJER1603).

References

- [1] G. L. Miller, “Riemann’s hypothesis and tests for primality,” *J. Comput. Syst. Sci.*, vol. 13, no. 3, pp. 300–317, 1976.
- [2] M. O. Rabin, “Probabilistic algorithm for testing primality,” *Journal of Number Theory*, vol. 12, no. 1, pp. 128 – 138, 1980.
- [3] M. Agrawal, N. Kayal, and N. Saxena, “PRIMES is in P,” *Ann. of Math*, vol. 2, pp. 781–793, 2002.
- [4] F. D. Wood, J. van de Meent, and V. Mansinghka, “A new approach to probabilistic programming inference,” in *Proc. of AISTATS 2014*, 2014, pp. 1024–1032.
- [5] N. D. Goodman, V. K. Mansinghka, D. M. Roy, K. Bonawitz, and J. B. Tenenbaum, “Church: a language for generative models,” in *Proc. of UAI 2008*, 2008, pp. 220–229.
- [6] C. Jones, “Probabilistic non-determinism,” Ph.D. dissertation, University of Edinburgh, 1990.
- [7] V. Danos and R. Harmer, “Probabilistic game semantics,” *ACM Trans. Comput. Log.*, vol. 3, no. 3, pp. 359–382, 2002.
- [8] T. Ehrhard, M. Pagani, and C. Tasson, “Full abstraction for probabilistic PCF,” *J. ACM*, vol. 65, no. 4, pp. 23:1–23:44, 2018.
- [9] U. Dal Lago and C. Grellois, “Probabilistic termination by monadic affine sized typing,” in *Proc. of ESOP 2017*, 2017, pp. 393–419.
- [10] F. Breuvert and U. Dal Lago, “On intersection types and probabilistic lambda calculi,” in *Proc. of PPDP 2018*, 2018, pp. 8:1–8:13.
- [11] U. Dal Lago, D. Sangiorgi, and M. Alberti, “On coinductive equivalences for higher-order probabilistic functional programs,” in *Proc. of POPL 2014*, 2014, pp. 297–308.
- [12] R. Crubillé and U. Dal Lago, “On probabilistic applicative bisimulation and call-by-value λ -calculi,” in *Proc. of ESOP 2014*, 2014, pp. 209–228.
- [13] T. Sato, A. Aguirre, G. Barthe, M. Gaboardi, D. Garg, and J. Hsu, “Formal verification of higher-order probabilistic programs: reasoning about approximation, convergence, bayesian inference, and optimization,” *PACMPL*, vol. 3, pp. 38:1–38:30, 2019.
- [14] A. Jung and R. Tix, “The troublesome probabilistic powerdomain,” *Electr. Notes Theor. Comput. Sci.*, vol. 13, pp. 70–91, 1998.
- [15] C. Heunen, O. Kammar, S. Staton, and H. Yang, “A convenient category for higher-order probability theory,” in *Proc. of LICS 2017*. IEEE Computer Society, 2017, pp. 1–12.
- [16] M. Vákár, O. Kammar, and S. Staton, “A domain theory for statistical probabilistic programming,” *Proc. ACM Program. Lang.*, vol. 3, no. POPL, pp. 36:1–36:29, 2019.
- [17] V. Danos and T. Ehrhard, “Probabilistic coherence spaces as a model of higher-order probabilistic computation,” *Inf. Comput.*, vol. 209, no. 6, pp. 966–991, 2011.
- [18] T. Ehrhard, M. Pagani, and C. Tasson, “Measurable cones and stable, measurable functions: a model for probabilistic higher-order programming,” *PACMPL*, vol. 2, no. POPL, pp. 59:1–59:28, 2018.
- [19] J. M. E. Hyland and C. L. Ong, “On full abstraction for PCF: i, ii, and III,” *Inf. Comput.*, vol. 163, no. 2, pp. 285–408, 2000.

- [20] S. Abramsky, R. Jagadeesan, and P. Malacaria, “Full abstraction for PCF,” *Inf. Comput.*, vol. 163, no. 2, pp. 409–470, 2000.
- [21] J.-Y. Girard, “Geometry of interaction 1: Interpretation of system F,” in *Logic Colloquium ’88*, ser. Studies in Logic and the Foundations of Mathematics, R. Ferro, C. Bonotto, S. Valentini, and A. Zanardo, Eds. Elsevier, 1989, vol. 127, pp. 221–260.
- [22] U. Dal Lago, C. Faggian, B. Valiron, and A. Yoshimizu, “The geometry of parallelism: classical, probabilistic, and quantum effects,” in *Proc. of POPL 2017*, 2017, pp. 833–845.
- [23] S. Castellan, P. Clairambault, H. Paquet, and G. Winskel, “The concurrent game semantics of probabilistic PCF,” in *Proc. of LICS 2018*, 2018, pp. 215–224.
- [24] P. Clairambault and H. Paquet, “Fully abstract models of the probabilistic lambda-calculus,” in *Proc. of CSL 2018*, 2018, pp. 16:1–16:17.
- [25] H. Paquet and G. Winskel, “Continuous probability distributions in concurrent games,” *Electr. Notes Theor. Comput. Sci.*, vol. 341, pp. 321–344, 2018.
- [26] A. Joyal, R. Street, and D. Verity, “Traced monoidal categories,” *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 119, no. 3, pp. 447–468, 1996.
- [27] S. Abramsky, E. Haghverdi, and P. Scott, “Geometry of interaction and linear combinatory algebras,” *Mathematical Structures in Comp. Sci.*, vol. 12, no. 5, pp. 625–665, oct 2002.
- [28] J. Borgström, U. Dal Lago, A. D. Gordon, and M. Szymczak, “A lambda-calculus foundation for universal probabilistic programming,” in *Proc. of ICFP 2016*, 2016, pp. 33–46.
- [29] N. Hoshino, K. Muroya, and I. Hasuo, “Memoryful geometry of interaction: From coalgebraic components to algebraic effects,” in *Proc. of CSL-LICS 2014*, 2014.
- [30] K. Muroya, N. Hoshino, and I. Hasuo, “Memoryful geometry of interaction ii: Recursion and adequacy,” in *Proc. of POPL 2016*, 2016, pp. 748–760.
- [31] P. Billingsley, *Probability and Measure*, 2nd ed. John Wiley and Sons, 1986.
- [32] S. Staton, “Commutative semantics for probabilistic programming,” in *Proc. of ESOP 2017*, 2017, pp. 855–879.
- [33] M. Wand, R. Culpepper, T. Giannakopoulos, and A. Cobb, “Contextual equivalence for a probabilistic language with continuous random variables and recursion,” *Proc. ACM Program. Lang.*, vol. 2, no. ICFP, pp. 87:1–87:30, 2018.
- [34] E. Moggi, “Notions of computation and monads,” *Information and Computation*, vol. 93, no. 1, pp. 55–92, 1991, selections from 1989 IEEE Symposium on Logic in Computer Science.
- [35] J.-Y. Girard, “Linear logic,” *Theoretical Computer Science*, vol. 50, no. 1, pp. 1–101, 1987.
- [36] M. Hyland and A. Schalk, “Glueing and orthogonality for models of linear logic,” *Theoretical Computer Science*, vol. 294, no. 1, pp. 183–231, 2003, category Theory and Computer Science.
- [37] Y. Lafont, *From proof nets to interaction nets*, ser. London Mathematical Society Lecture Note Series. Cambridge University Press, 1995, pp. 225–248.
- [38] P. Selinger, *A Survey of Graphical Languages for Monoidal Categories*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 289–355.
- [39] O. Laurent, “A token machine for full geometry of interaction,” in *Proc. of TLCA 2001*, 2001, pp. 283–297.

- [40] S. Abramsky and G. McCusker, “Linearity, sharing and state: a fully abstract game semantics for idealized algol with active expressions: Extended abstract,” *Electronic Notes in Theoretical Computer Science*, vol. 3, pp. 2–14, 1996.
- [41] I. Mackie, “The geometry of interaction machine,” in *Proc. of POPL 1995*, 1995, pp. 198–208.
- [42] B. Jacobs, *Introduction to Coalgebra: Towards Mathematics of States and Observation*, ser. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2016.
- [43] U. Dal Lago and N. Hoshino, “The geometry of bayesian programming,” available at <https://arxiv.org/abs/1904.07425>.
- [44] G. Kelly and M. Laplaza, “Coherence for compact closed categories,” *Journal of Pure and Applied Algebra*, vol. 19, pp. 193–213, 1980.
- [45] S. M. Lane, *Categories for the Working Mathematician*, 2nd ed. Springer, 1998.
- [46] M. Hasegawa, “The uniformity principle on traced monoidal categories,” *Electronic Notes in Theoretical Computer Science*, vol. 69, pp. 137–155, 2003, cTCS’02, Category Theory and Computer Science.
- [47] N. Hoshino, K. Muroya, and I. Hasuo, “Memoryful geometry of interaction: From coalgebraic components to algebraic effects,” available at <http://www.kurims.kyoto-u.ac.jp/~naophiko/paper/ResGoI.pdf>.
- [48] K. Cho and B. Jacobs, “Disintegration and bayesian inversion, both abstractly and concretely,” 2017, to appear in *Mathematical Structures in Computer Science*.
- [49] B. Jacobs, F. Zanasi, and A. Kissinger, “Causal inference by string diagram surgery,” 2018, arXiv:1811.08338 [cs.LO].
- [50] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [51] W. K. Hastings, “Monte carlo sampling methods using markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.